

; Entered [jdelaval 2-Mar-05 12:07]
SG01
wxhpgf991

; Entered [jdelaval 2-Mar-05 12:08]
SG02
wxhpgfex1

Wed Mar 2 14:26:43 2005

! Entered [jdelavai 2-Mar-05 12:08]
SEO3
wxhpgfer1

robinson-10-664359.pcp

! Entered [jdelavai 2-Mar-05 12:08]
SEO4
kxhpgf991

Wed Mar 2 14:26:43 2005

robinson-10-664359 pep

Page 3

: Entered [jdelaval 2-Mar-05 12:09]
S805
kxhpgfek1

: Entered [jdelaval 2-Mar-05 12:09]
S806
kxhpgfer1

Wed Mar 2 14:26:43 2005

! Entered [jdelaval 2-Mar-05 12:10]
SEQ7
rxhpqf991

robinson-10-664359.pcp

! Entered [jdelaval 2-Mar-05 12:10]
SEQ8
rxhpqfek1

/ Entered [jdelaval 2-Mar-05 12:10]
SEQ9
rxhpgfer1

/ Entered [jdelaval 2-Mar-05 12:11]
SEQ10
wshpgfekyggswshpgfek1

Wed Mar 2 14:26:43 2005

robinson-10-664359.pgp

Page 6

! Entered [jdelaval 2-Mar-05 12:11]
SEQ11
wshpqfeksgsgsgswshpqfek1

! Entered [jdelaval 2-Mar-05 12:12]
SEQ12
wshpqfekxxxxxxxxwshpqfek1

/ Entered [jdelaval 2-Mar-05 12:12]
SEQ13
wshpqfexxxxxxxxxxwshpqfex1

/ Entered [jdelaval 2-Mar-05 12:12]
SEQ14
wshpqfexxxxxxxxxxwshpqfex1

; Entered [jdelaval 2-Mar-05 12:12]
SEQ15
wshpqfexxxxxxxxxxwshpqfex1

; Entered [jdelaval 2-Mar-05 12:12]
SEQ16
wshpqfexxxxxxxxxxwshpqfex1

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using 6w model

Run on: March 2, 2005, 12:20:44 ; Search time 6.11382 Seconds

(without alignments)
125.901 Million cell updates/sec

Title: SEQ1

Perfect score: 50

Sequence: 1 wxhpgfgg 8

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR79:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	50	100.0	131	2 S66537	Ig heavy chain V r
2	45	90.0	452	2 P83587	inner membrane pro
3	41	82.0	363	2 T26842	hypothetical prote
4	40	80.0	288	2 T10477	sec3 protein - ye
5	40	80.0	297	2 T13905	protein transport
6	40	80.0	297	2 A45442	transport vesicle
7	40	80.0	512	2 H86216	protein T3G18.16
8	39	78.0	224	2 AF3382	alpha/beta hydrola
9	39	78.0	225	2 B97580	hypothetical prote
10	39	78.0	225	2 AH2800	conserved hypothet
11	39	78.0	339	2 T37727	probable nuclear p
12	39	78.0	436	1 S08156	gene T protein - m
13	38	76.0	396	1 UGECM	periplasmic maltos
14	38	76.0	396	2 AD1014	periplasmic maltos
15	38	76.0	396	2 A98256	periplasmic maltos
16	38	76.0	396	2 E86096	periplasmic maltos
17	38	76.0	396	2 S05331	maltose-binding pr
18	38	76.0	403	2 AB0452	maltose-binding pe
19	38	76.0	406	2 H82397	maltose ABC transp
20	38	76.0	471	2 E87340	M20/M25/M40 family
21	38	76.0	587	2 B83324	hypothetical prote
22	37	74.0	302	2 T02480	sec3-related prot
23	37	74.0	535	2 B83258	alkaline serine pr
24	37	74.0	685	2 H83428	conserved hypothet
25	37	74.0	821	2 A12417	hypothetical prote
26	37	74.0	1245	2 H83574	conserved hypothet
27	37	74.0	3149	1 Q08B8	BP1P1 protein - hu
28	36	72.0	85	2 A05126	hypothetical prote
29	36	72.0	126	2 I44151	Ig heavy chain V r

30	36	72.0	318	1 DB8SLM	L-lactate dehydrog
31	36	72.0	338	2 S74042	probable pyruvate
32	36	72.0	362	2 C90674	hypothetical prote
33	36	72.0	362	2 C64758	Yanb protein - Esc
34	36	72.0	365	2 F85524	hypothetical prote
35	36	72.0	449	2 AF1075	inner membrane pro
36	36	72.0	471	2 AG0146	probable pyridoxal
37	36	72.0	492	2 AG0109	probable membrane
38	36	72.0	1001	2 T17365	serine/threonine p
39	35	70.0	116	2 A95862	hypothetical prote
40	35	70.0	137	2 T17240	hypothetical prote
41	35	70.0	238	2 C83699	necrosis and ethyl
42	35	70.0	258	2 F71818	hypothetical prote
43	35	70.0	264	2 C64701	hypothetical prote
44	35	70.0	288	2 T12462	hypothetical prote
45	35	70.0	365	2 AB1272	X-Pro dipeptidase

ALIGNMENTS

```
RESULT 1
S66537
Ig heavy chain V region - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: S66537
R:Tsiclotis, G.; Haase, W.; Engel, A.; Michel, H.
Eur. J. Biochem. 231, 823-830, 1995
A:Title: Isolation and structural characterization of trimeric cyanobacterial photosyste
A:Reference number: S66536; MUID:95377318; PMID:7649183
A:Accession: S66537
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-131 <TSI>
A:Cross-references: EMBL:X88902; NID:9895869; PIDN:CAA61364.1; PID:g1103701
C:Superfamily: immunoglobulin V region: immunoglobulin homology
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match          100.0%; Score 50; DB 2; Length 131;
Best Local Similarity 87.5%; Pred. No. 0.14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WYHPQFG 8
DB 124 WRHPQFG 131

RESULT 2
P83587
Inner membrane protein Cred PA0465 (imported) - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 12-Jul-2004
A:Accession: P83587
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: P83587
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-452 <STO>
A:Cross-references: UNIPROT:Q91653; GB:AE004484; GB:AE004091; NID:g9946320; PIDN:AA0385
A:Experimental source: strain PA01
C:Genetics:
A:Gene: cred; PA0465
C:Superfamily: inner membrane protein Cred

Query Match          90.0%; Score 45; DB 2; Length 452;
Best Local Similarity 75.0%; Pred. No. 3.6;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

QY 1 WKHPQFG 8
Db 234 WHPSPFG 241

RESULT 3

126842
hypothetical protein Y43F4B.4 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 16-Aug-2004
R:Matthews, L.
submitted to the EMBL Data Library, January 1998
A:Reference number: 220276
A:Accession: T26842
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-363 <Wtl>
A:Cross-references: UNIPROT:O45933; EMBL:AL021481; PIDN:CAA16333.1; CESP:Y43F4B.4
C:Genetic: CESP:Y43F4B.4
A:Gene: CESP:Y43F4B.4
A:Introns: 22/3; 59/3; 80/2; 103/1; 232/2; 318/1
C:Superfamily: WD repeat homology

Query Match
Best Local Similarity 82.0%; Score 41; DB 2; Length 363;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFG 7
Db 70 WAHPKFG 76

RESULT 4

110477
sec13 protein - yeast (*Pichia pastoris*)
C:Species: *Pichia pastoris*
C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Aug-2004
R:Payne, W.E.; Kaiser, C.A.
submitted to the EMBL Data Library, March 1996
A:Description: Isolation of genes encoding components of the secretory pathway from *Pichia*
A:Reference number: 217041
A:Accession: T10477
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-288 <PAY>
A:Cross-references: UNIPROT:P53024; EMBL:U52430; NID:g1277235; PID:g1277236
C:Genetic: C:Genetic: SGC13
A:Gene: SGC13
C:Function:
A:Description: required for the formation of mature transport vesicles from the ER membr
C:Superfamily: WD repeat homology
C:Keywords: peripheral membrane protein

Query Match
Best Local Similarity 80.0%; Score 40; DB 2; Length 288;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFG 7
Db 60 WAHPKFG 66

RESULT 5

139905
protein transport protein sec13 homolog, WD domain containing - fission yeast (*Schizosac*
C:Species: *Schizosaccharomyces pombe*
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Aug-2004
C:Accession: T39905

R:Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, November 1998
A:Reference number: 221889
A:Accession: T39905
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-297 <LYN>
A:Cross-references: UNIPROT:O94319; EMBL:AL033534; PIDN:CAA22129.1; GSPDB:GN00067; SPDB
C:Genetic: SGPDB:SPBC215.15
A:Gene: SGPDB:SPBC215.15
A:Map position: 2
A:Introns: 12/3
C:Superfamily: WD repeat homology

Query Match
Best Local Similarity 80.0%; Score 40; DB 2; Length 297;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFG 7
Db 61 WAHPKFG 67

RESULT 6

A45442
transport vesicle formation protein SEC13 [validated] - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: protein L8167.4; protein YLR208W
C:Species: *Saccharomyces cerevisiae*
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
R:Przyer, N.K.; Salama, N.R.; Schekman, R.; Kaiser, C.A.
J. Cell Biol. 120, 865-875, 1993
A:Title: Cytosolic Sec13p complex is required for vesicle formation from the endoplasm
A:Reference number: A45442; MUID:93163112; PMID:8432727
A:Molecule type: DNA
A:Residues: 1-297 <PRY>
A:Cross-references: UNIPROT:O04491; EMBL:L05929; NID:g172558; PID:g172559
A:Note: sequence extracted from NCBI backbone (NCBIN:124845, NCBI:P:124846)
R:Pauley, A.
submitted to the EMBL Data Library, September 1994
A:Description: The sequence of *S. cerevisiae* cosmid 8167.
A:Reference number: S48545
A:Accession: S48559
A:Molecule type: DNA
A:Residues: 1-297 <PAU>
A:Cross-references: EMBL:U14913; NID:g544497; PID:g544501; GSPDB:GN00012; MIPS:YLR208w
R:Simosoglou, S.; Wimmer, C.; Rieger, M.; Doye, V.; Tekotte, H.; Weise, C.; Emig, S.;
Cell 84, 265-275, 1996
A:Title: A novel complex of nucleoporins, which includes Sec13p and a Sec13p homolog, is
A:Reference number: S62137; MUID:96152656; PMID:8565072
A:Accession: S62146
A:Molecule type: protein
A:Residues: 278-282; 125-131; 79-83; 35-39; 289-293 <SIN>
C:Genetic: SGC13
A:Gene: SGD:SGC13; MIPS:YLR208w
A:Cross-references: SGD:S0004198; MIPS:YLR208w
A:Map position: 12R
C:Function:
A:Description: required for the formation of mature transport vesicles from the ER membr
C:Superfamily: unassigned WD repeat proteins; WD repeat homology
C:Keywords: peripheral membrane protein

Query Match
Best Local Similarity 80.0%; Score 40; DB 2; Length 297;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFG 7
Db 61 WAHPKFG 67

```

RESULT 7
H66216
protein T23G18.16 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: H66216
R:Thellogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewart, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Author: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzilli,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Author: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Ye, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A66141; MUID:21016719; PMID:11130712
A:Accession: H66216
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-512 <STO>
A:Cross-references: UNIPROT:Q9SGD0; GB:AE005172; NID:g6579199; PIDN:AAF18242.1; GSPDB:GN
C:Genetics:
A:Gene: T23G18.16
A:Map position: 1

Query Match
Best Local Similarity 80.0%; Score 40; DB 2; Length 512;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XHPQFG 8
Db 111 PHPQFG 118

RESULT 8
AF3382
alpha/beta hydrolase [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AF3382
R:DeiVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
Mazur, M.; Goldtman, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Leles
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: A03252; PMID:11756688
A:Accession: AF3382
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-224 <KUR>
A:Cross-references: UNIPROT:Q8YGM1; UNIPROT:Q8G102; GB:AE008917; PIDN:AAU52225.1; PID:gl
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME1044
A:Map position: 1

Query Match
Best Local Similarity 78.0%; Score 39; DB 2; Length 224;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
Db 33 PHPQFG 39

RESULT 9
B97580
hypothetical protein AGR_C_3351 [imported] - Agrobacterium tumefaciens (strain C58, Cer
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C:Accession: B97580
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;

```

```

Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tun
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: B97580
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-225 <KUR>
A:Cross-references: UNIPROT:Q8UBD4; GB:AE007669; PIDN:AAK87595.1; PID:gl5156937; GSPDB:GN
C:Genetics:
A:Gene: AGR_C_3351
A:Map position: circular chromosome

Query Match
Best Local Similarity 78.0%; Score 39; DB 2; Length 225;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
Db 33 PHPQFG 39

RESULT 10
AH2800
conserved hypothetical protein Atu1826 [imported] - Agrobacterium tumefaciens (strain C5
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C:Accession: AH2800
R:Wood, D.W.; Setudal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavhin, T.; Levy, R.; Li, M.; McClell
i Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Author: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AH2800
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-225 <KUR>
A:Cross-references: UNIPROT:Q8UBD4; GB:AE008668; PIDN:AAU42822.1; PID:gl7740269; GSPDB:GN
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu1826
A:Map position: circular chromosome

Query Match
Best Local Similarity 78.0%; Score 39; DB 2; Length 225;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
Db 33 PHPQFG 39

RESULT 11
T37727
probable nuclear pore protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 16-Aug-2004
C:Accession: T37727
R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z21740
A:Accession: T37727
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Cross-references: UNIPROT:Q10099; EMBL:268197; PIDN:CAA92379.1; GSPDB:GN00066; SPDB:SP
A:Experimental source: strain 972h-; cosmid c15F9
C:Genetics:
A:Gene: SPDB:SPAC15F9.02
A:Map position: 1
A:introns: 23/1, 334/2

```

C:Superfamily: WD repeat homology

Query Match

Best Local Similarity 78.0%; Score 39; DB 2; Length 339;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXPQFG 7
|:|:|
68 WAHPSFG 74

RESULT 12

S08156

Gene T protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

R:Accession: S08156

R:Herzmann, B.G.; Label: S.; Pousetka, A.; King, T.R.; Lehrach, H.

A:Title: Cloning of the T gene required in mesoderm formation in the mouse.

A:Reference number: S08156; MUID:90158787; PMID:2154694

A:Accession: S08156

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-436 <HR>

A:Cross-references: UNIPROT:P20293; GB:X51683; NID:955053; PID:CAA35985.1; PID:G55054

A:Gene: T

A:Map position: 17

C:Superfamily: mouse gene T protein; T-box homology

C:Keywords: DNA binding; transcription regulation

F:1-227/Domain: T-box homology <TBX>

Query Match

Best Local Similarity 78.0%; Score 39; DB 1; Length 436;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
|:|:|
260 SHPQFG 266

RESULT 13

JGECM

periplasmic maltose-binding protein precursor - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 09-Jul-2004

R:Accession: A03428; I54874; A65211; I54911

R:Display, P.; Bedouelle, H.; Fowler, A.; Zabin, I.; Saurin, W.; Hofnung, M.

A:Title: Sequences of the maltose gene and of its product, the maltose-binding protein of E.

A:Reference number: A03428; MUID:84289460; PMID:6088507

A:Accession: A03428

A:Molecule type: DNA

A:Residues: 1-396 <DUP>

A:Cross-references: UNIPROT:P02928; GB:J01648; GB:J01639; GB:K02117; GB:M24344; GB:M2434

A:Experimental source: strain K12

A:Note: Most of the primary structure was confirmed by protein sequencing

J. Bacteriol. 169, 1794-1800, 1987

A:Title: Suppression of a signal sequence mutation by an amino acid substitution in the m

A:Reference number: I54874; MUID:87194547; PMID:3553148

A:Accession: I54874

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-28 <RES>

A:Cross-references: GB:M16181; NID:9146704; PID:AAA24102.1; PID:9146705

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Accession: A65211

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-396 <BLAT>

A:Cross-references: GB:A000476; GB:U00096; NID:91790456; PID:AACT7004.1; PID:91790466;

A:Experimental source: strain K-12, substrain M61655

R:Collier, D.N.; Bassford, P.J.

J. Bacteriol. 171, 4640-4647, 1989

A:Title: Mutations that improve export of maltose-binding protein in SecB- cells of Es.

A:Reference number: I54911; MUID:89359092; PMID:2670890

A:Accession: I54911

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-15; 'K', 17-29 <RE2>

A:Cross-references: GB:M29860; NID:9146885; PID:AAA24179.1; PID:9146886

C:Comment: This periplasmic binding protein is involved in the high-affinity maltose m-

C:Genetics:

A:Gene: malt

A:Map position: 92 min

C:Superfamily: maltose-binding protein

C:Keywords: periplasmic space; sugar transport

F:1-26/Domain: signal sequence #status predicted <SIG>

F:27-396/Product: maltose-binding protein #status experimental <MAT>

Query Match
Best Local Similarity 76.0%; Score 38; DB 1; Length 396;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXPQFG 8
|:|:|
88 WAHDPFG 95

RESULT 14

AD1014

periplasmic maltose-binding protein [imported] - Salmonella enterica subsp. enterica ser.

C:Species: Salmonella enterica subsp. enterica serovar Typh

A:Note: this species has also been called Salmonella typh

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002

R:Parikh, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; MUID:21534947; PMID:11677608

A:Accession: AD1014

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-396 <PAR>

A:Cross-references: GB:ML513382; PIDN:CAD09213.1; PID:g16505217; GSPDB:GN00176

C:Genetics:

A:Gene: STY4425

C:Superfamily: maltose-binding protein

Query Match

Best Local Similarity 76.0%; Score 38; DB 2; Length 396;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXPQFG 8
|:|:|
88 WAHDPFG 95

RESULT 15

A98256

periplasmic maltose-binding protein precursor - Escherichia coli (strain O157:H7, subser

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004

C:Accession: A98256

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gaewara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A/Reference number: A96629; MUID:21156231; PMID:1128796
A/Accession: A96256
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-396 <HAY>
A/Cross-references: UNIPROT:P02928; GB:BA000007; PION:BA838440.1; PID:gl3364494; GSPDB:C
A/Experimental source: strain O157:H7, substrain RMD 0509952
A/Genetics:
C/Gene: EC85017
C/Superfamily: maltose-binding protein

Query Match 76.0%; Score 38; DB 2; Length 396;
Best Local Similarity 62.5%; Pred. No. 50;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WXHPOFGG 8
|:|:|
Db 88 WAHDRFGG 95

Search completed: March 2, 2005, 12:28:47
Job time : 8.11382 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 29.7236 Seconds
(without alignments)
137.824 Million cell updates/sec

Title: SEQ1
Perfect score: 50
Sequence: 1 wxhpqfsg 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	90.0	439	2	Q8P7B9
2	45	90.0	439	2	Q8P7B9
3	45	90.0	439	2	Q8P7B9
4	44	88.0	127	2	Q65QC7
5	44	88.0	620	2	Q82CV1
6	44	88.0	634	2	Q93J80
7	43	86.0	211	2	Q9N4A6
8	43	86.0	213	2	Q9N4A7
9	43	86.0	474	2	Q80YU0
10	42	84.0	257	2	Q6R560
11	41	82.0	186	2	Q7ZU38
12	41	82.0	259	1	PRI0.TRIUV
13	41	82.0	295	2	Q9APW1
14	41	82.0	295	2	Q88J42
15	41	82.0	354	2	Q7KQ53
16	41	82.0	357	2	Q8R2U0
17	41	82.0	360	2	Q6GNF1
18	41	82.0	363	2	Q45933
19	41	82.0	364	2	Q6RGU2
20	41	82.0	381	2	Q7O9L2
21	41	82.0	401	2	Q7NV41
22	41	82.0	411	2	Q9N4A8
23	41	82.0	421	1	SEH1.HUMAN
24	41	82.0	658	2	Q8BE51
25	41	82.0	729	2	Q6YV06
26	41	82.0	754	2	Q88I20
27	41	82.0	811	2	Q877H7
28	41	82.0	2377	2	Q7TP06
29	41	82.0	289	1	SC13.PICPA
30	40	80.0	295	2	Q75B52
31	40	80.0	295	2	Q75B52

32	40	80.0	297	1	SC13.YEAST	Q04491.saccharomyc
33	40	80.0	297	2	Q94319	Q94319.schizosacch
34	40	80.0	297	2	Q6B1R1	Q6B1R1.debaromyce
35	40	80.0	298	2	Q6B2X5	Q6B2X5.yarrowia_11
36	40	80.0	298	2	Q6FNV4	Q6FNV4.candida_gla
37	40	80.0	301	2	Q81AS2	Q81AS2.arabidopsis
38	40	80.0	301	2	Q81IA2	Q81IA2.oryza_sativ
39	40	80.0	302	2	Q6CSZ5	Q6CSZ5.kluyveromyc
40	40	80.0	302	2	Q6SR11	Q6SR11.arabidopsis
41	40	80.0	356	2	Q9V3J4	Q9V3J4.drosophila
42	40	80.0	386	2	Q7KLW8	Q7KLW8.drosophila
43	40	80.0	512	2	Q9SGD0	Q9SGD0.arabidopsis
44	40	80.0	562	2	Q8EAP2	Q8EAP2.shewanella
45	40	80.0	784	2	Q87XH0	Q87XH0.pseudomonas

ALIGNMENTS

RESULT 1
ID Q8P7B9 PRELIMINARY; PRT; 439 AA.
AC Q8P7B9;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Inner membrane protein.
GN Name=cred; OrderedLocNames=XCC2692;
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OK NCBI_Taxid=340;
RN [1]
RP SEQUENCE FROM N.A. / NCPB 528;
RC STRAIN=ATCC 33913 / NCPIB 528;
RX MEDLINE=22022145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camnayan F., Cardozo J., Chamergo F., Clapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fontighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katayama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Wenck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing host specificities.";
RT Nature 417:459-463 (2002).
DR EMBL: AE012381; AAM41964.1; -;
DR GO: GO:000175; F.3'-5'-exoribonuclease activity; IEA.
DR GO: GO:0003723; F:RNA binding; IEA.
DR GO: GO:0006356; P:RNA processing; IEA.
DR InterPro: IPR010247; 3_Exonase.
DR InterPro: IPR010364; Cred.
DR Pfam: PF06123; Cred; 1.
KW Complete proteome.
SQ SEQUENCE 439 AA; 48001 MW; A6F60F8090E7394 CRC64;

Query Match 90.0%; Score 45; DB 2; Length 439;
Best Local Similarity 75.0%; Pred. No. 19;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 WYHPQFG 8
DB 235 WQHPFGG 242

```

RESULT 2
ID Q8PIP3 PRELIMINARY; PRT; 439 AA.
AC Q8PIP3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DE Inner membrane protein.
GN Name=Cred; OrderedLocNames=XAC2852;
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
NCBI_TaxID=92829;
XN Xanthomonadaceae; Xanthomonas.
RN [1]
RP STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=2022145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Queglio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camayan F., Cardozo J., Chambergro F., Chapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.P.,
RA Spínola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463(2002).
DR EMBL; AB011926; AAM37697.1;
DR InterPro; IPRO10364; Cred.
DR Pfam; PF06123; Cred; 1.
KW Complete proteome.
SQ SEQUENCE 439 AA; 48419 MW; 5FC3D2C7AC8F4FC CRC64;

Query Match
Best Local Similarity 90.0%; Score 45; DB 2; Length 439;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXPQFGG 8
Db 235 WXPQFGG 242

RESULT 3
ID Q91653 PRELIMINARY; PRT; 452 AA.
AC Q91653;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE Inner membrane protein.
GN Name=Cred; OrderedLocNames=PA0465;
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
NCBI_TaxID=287;
XN Pseudomonadaceae; Pseudomonas.
RN [1]
RP STRAIN=ATCC 15692 / PA01;
RX MEDLINE=2043737; PubMed=10984043; DOI=10.1038/35023079;
RA Stever C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.U., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrock-Madman S., Yuan Y.,
RA Brody L., Coulter S.N., Folger K.R., Kas A., Lardiz G., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Retzer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;

```

```

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL; AB004484; AAG03854.1;
DR PIR; F83587; F83587.
DR InterPro; IPRO10364; Cred.
DR Pfam; PF06123; Cred; 1.
KW Complete proteome.
SQ SEQUENCE 452 AA; 50550 MW; B2E8BF375A7F975C CRC64;

Query Match
Best Local Similarity 90.0%; Score 45; DB 2; Length 452;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXPQFGG 8
Db 234 WXPQFGG 241

RESULT 4
ID Q650C7 PRELIMINARY; PRT; 127 AA.
AC Q650C7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DE Crib protein.
GN Name=Cred; ORFNames=MS2226;
OS Mannheimia succiniciproducens MBE155R.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Mannheimia.
NCBI_TaxID=221988;
RN [1]
RP STRAIN=MBE155R.
RX STRAIN=MBE155R.
RA Hong S.H., Kim J.S., Lee S.Y., In Y.H., Choi S.S., Rih J.-K.,
RA Kim C.H., Jeong H., Hur C.-G., Kim J.G.;
RT "The genome sequence of the capnophilic rumen bacterium Mannheimia
RT succiniciproducens."
RL Nat. Biotechnol. 0:0-0(2004).
DR EMBL; AB016827; AAU38833.1;
SQ SEQUENCE 127 AA; 14089 MW; B844F39F46C38B8C CRC64;

Query Match
Best Local Similarity 88.0%; Score 44; DB 2; Length 127;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXPQFGG 8
Db 58 WXPQFGG 65

RESULT 5
ID Q82CV1 PRELIMINARY; PRT; 620 AA.
AC Q82CV1;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Putative sensory box/GDPF family protein.
GN OrderedLocNames=SAV537;
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycetales; Streptomycetaceae; Streptomyces.
NCBI_TaxID=33903;
RN [1]
RP STRAIN=MA-4680;
RX MEDLINE=21477403; PubMed=11572948; DOI=10.1073/pnas.211433198;
RA Omura S., Ikeda H., Itohikawa J., Hanamoto A., Takahashi C.,
RA Shinoue M., Takahashi Y., Horikawa H., Nakazawa H., Osorio T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces

```

```

RT avermitilis: deducing the ability of producing secondary
RT merbolites.";
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680;
RX MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis.";
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005042; BAC72949.1;
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0007165; P:signal transduction; IEA.
DR InterPro; IPR001633; EAL.
DR InterPro; IPR001633; EAL.
DR InterPro; IPR001610; GDEF.
DR InterPro; IPR001610; PAC.
DR Pfam; PF00563; EAL; 1.
DR Pfam; PF00990; GDEF; 1.
DR SMART; SM00267; DUF1; 1.
DR SMART; SM00052; DUF2; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 1.
DR TIGRfam; TIGR00254; GDEF; 1.
DR TIGRfam; TIGR00229; sensory_box; 1.
DR PROSITE; PS50883; EAL; 1.
DR PROSITE; PS50887; GDEF; 1.
DR PROSITE; PS50112; PAS; 1.
KW Complete proteome.
SQ SEQUENCE 620 AA; 67211 MW; 86D39D607B8A42ED CRC64;

```

```

Query Match      88.0%; Score 44; DB 2; Length 620;
Best Local Similarity 85.7%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 WXPQFG 7
Db 363 WNPQFG 389

```

```

RESULT 6
Q93J80 PRELIMINARY; PRT; 634 AA.
AC Q93J80;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein SC02817.
GN ORFNames=SCBAC17F8.08c;
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=1902;
RN
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=1200953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieffer H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieffer T., Larke L., Murphy L.D., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
EMBL; AL939114; CAC44280.1; -.

```

```

DR GO; GO:0000155; F:two-component sensor molecule activity; IEA.
DR GO; GO:0000160; P:two-component signal transduction system (p. . .; IEA.
DR InterPro; IPR001633; EAL.
DR InterPro; IPR001633; EAL.
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS.
DR InterPro; IPR000700; PAS-associ_C.
DR Pfam; PF00563; EAL; 1.
DR Pfam; PF00990; GDEF; 1.
DR SMART; SM00267; DUF1; 1.
DR SMART; SM00052; DUF2; 1.
DR SMART; SM00086; PAC; 1.
DR SMART; SM00091; PAS; 1.
DR TIGRfam; TIGR00254; GDEF; 1.
DR TIGRfam; TIGR00229; sensory_box; 1.
DR PROSITE; PS50883; EAL; 1.
DR PROSITE; PS50887; GDEF; 1.
DR PROSITE; PS50113; PAC; 1.
DR PROSITE; PS50112; PAS; 1.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 634 AA; 67196 MW; B53425D330A29526 CRC64;

```

```

Query Match      88.0%; Score 44; DB 2; Length 634;
Best Local Similarity 85.7%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 WXPQFG 7
Db 390 WNPQFG 396

```

```

RESULT 7
Q9N4A6 PRELIMINARY; PRT; 211 AA.
AC Q9N4A6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Nuclear pore complex protein protein 20, isoform b.
GN Name=pp-20; ORFNames=Y77E11A.13;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RG WormBase Consortium;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Dante M.;
RT "The sequence of C. elegans cosmid Y77E11A.";
RN
RN [3]
RP Submitted (Mar-2000) to the EMBL/GenBank/DBJ databases.
RN
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=11996410; PubMed=1200953; DOI=10.1038/417141a;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieffer H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieffer T., Larke L., Murphy L.D., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
EMBL; AL939114; CAC44280.1; -.

```



```

ID  Prio Trivu  STANDARD;  PRT;  259 AA.
AC  P51780;
DT  01-OCT-1996 (Rel. 34, Created)
DT  01-OCT-1996 (Rel. 34, Last sequence update)
DE  05-JUL-2004 (Rel. 44, Last annotation update)
DE  Major prion protein precursor (PrP) (P27-30) (PrP33-35C).
GN  Name=PrNP;
OS  Trichosturus vulpecula (Brush-tailed possum).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosturus.
ON  NCBI_TaxID=9337;
[1]
RP  SEQUENCE FROM N.A.
RA  Windl O., Dempster M., Escibeiro P., Lathe R.;
RL  Submitted (MAY-1995) to the EMBL/GenBank/DBD databases.
CC  - FUNCTION: The function of PrP is not known. PrP is encoded in the
CC  host genome and is expressed both in normal and infected cells.
CC  - SUBUNIT: PrP has a tendency to aggregate yielding polymers called
CC  "rods".
CC  - SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC  - DISEASE: PrP is found in high quantity in the brain of humans and
CC  animals infected with the degenerative neurological diseases kuru,
CC  Creutzfeldt-Jakob disease (CJD), Gerstmann-Strausler syndrome
CC  (GSS), scrapie, bovine spongiform encephalopathy (BSE),
CC  transmissible mink encephalopathy (TME), etc.
CC  - SIMILARITY: Belongs to the prion family.
-----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
-----
DR  HMBL; L38993; AAA61833.1; -
DR  HSSP; P10279; IDWY.
DR  InterPro; IPR000817; Prion.
DR  Pfam; PF00377; Prion; 1.
DR  PRINTS; PR00341; PRION.
DR  PROSITE; PS00291; PRION_1; 1.
DR  PROSITE; PS00706; PRION_2; 1.
KW  Glycoprotein; GPI-anchor; Lipoprotein; Membrane; Prion; Repeat;
FT  SIGNAL.
FT  CHAIN.
FT  PROPEP.
FT  DISULFID.
FT  LIPID.
FT  CARBOHYD.
FT  CARBOHYD.
SQ  SEQUENCE 259 AA; 28550 MW; 27DA7DBEBCA56892 CRC64;

Query Match
Best Local Similarity 82.0%; Score 41; DB 1; Length 259;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY  1 WXPDPFG 8
Db  58 WXPDPFG 65

RESULT 13
ID  Q9APW1  PRELIMINARY;  PRT;  295 AA.
AC  Q9APW1;
DT  01-JUN-2001 (TrEMBLrel. 17, Created)
DT  01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Hypothetical protein.
OS  Pseudomonas aeruginosa.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

```

```

OC  Pseudomonadaceae; Pseudomonas.
ON  NCBI_TaxID=287;
[1]
RP  SEQUENCE FROM N.A.
RA  Liang X., Pham X.Q.T., Olson M.V., Lory S.;
RL  "Identification of a genomic island present in the majority of
RL  pathogenic isolates of Pseudomonas aeruginosa."
RL  U. Bacteriol. 183:843-853(2001).
DR  EMBL; AF241171; AA01512.1; -
DR  InterPro; IPR008775; PhyH.
DR  Pfam; PF05721; PhyH; 1.
KW  Hypothetical protein.
SQ  SEQUENCE 295 AA; 32712 MW; 0155D498316202D5 CRC64;

Query Match
Best Local Similarity 82.0%; Score 41; DB 2; Length 295;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY  1 WXPDPFG 7
Db  143 WXPDPFG 149

RESULT 14
ID  Q88J42  PRELIMINARY;  PRT;  295 AA.
AC  Q88J42;
DT  01-JUN-2003 (TrEMBLrel. 24, Created)
DT  01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Hypothetical protein.
GN  OrderedLocustNames=PP2807;
OS  Pseudomonas putida (strain KT2440).
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC  Pseudomonadaceae; Pseudomonas.
ON  NCBI_TaxID=160486;
[1]
RP  SEQUENCE FROM N.A.
RA  MEDLINE=22423060; PubMed=12534463;
RA  Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA  Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA  Brinkac L.M., Beaman M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,
RA  Madupu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,
RA  Hance I., Chris Lee P., Holtzapfe E.K., Scanlan D., Tran K.,
RA  Moazed A., Uetebach T.R., Rizzo M., Lee K., Kosack D., Moestl D.,
RA  Medler H., Lauber J., Stjepandic D., Hohenel J., Straetz M., Heim S.,
RA  Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tummeler B.,
RA  Frazer C.M.;
RT  "Complete genome sequence and comparative analysis of the
RT  metabolically versatile Pseudomonas putida KT2440."
RL  Environ. Microbiol. 4:799-808(2002).
DR  EMBL; AE016784; AAN68415.1; -
DR  TIGR; PP2807; -
DR  InterPro; IPR008775; PhyH.
DR  Pfam; PF05721; PhyH; 1.
KW  Complete proteome.
SQ  SEQUENCE 295 AA; 32391 MW; EBBE0A701DF647 CRC64;

Query Match
Best Local Similarity 82.0%; Score 41; DB 2; Length 295;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY  1 WXPDPFG 7
Db  143 WXPDPFG 149

RESULT 15
ID  Q7K0S3  PRELIMINARY;  PRT;  354 AA.
AC  Q7K0S3;
DT  05-JUL-2004 (TrEMBLrel. 27, Created)

```


DT 05-JUL-2004 (TrEMBLrel. 27, last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, last annotation update)
 DE CG8722-PA (CG8722-pb) (SD07614p).
 GN Name=Nup44a; Synonym=nup44a; ORFNames=CG8722;
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 CC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Ephydroidea; Drosophilidae; Drosophila.
 CX NCBI_taxid=7227;
 OK
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
 RA Adams M.D., Celnikner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabore G.L.,
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Boriva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durkin K.J., Evangelista C.C., Ferraz S., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Jeggam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kension J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Liao X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mactel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mervulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spelling A.C., Stapleton M., Strong R., Smith T.,
 RA Svrtkac R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Wesserman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 RL
 RN
 RN SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537573;
 RA Celnikner S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
 RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svrtkac R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
 RT melanogaster euchromatic genome sequence.";
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RL
 RN
 RN SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svrtkac R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celnikner S.E.;
 RT "The transposable elements of the Drosophila melanogaster euchromatin:
 RT a genome perspective.";
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RL
 RN
 RN SEQUENCE FROM N.A.

RX MEDLINE=22426069; PubMed=12537572;
 RA Mista S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Belencourt B.R., Celnikner S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a
 RT systematic review.";
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RL
 RN
 RN SEQUENCE FROM N.A.
 RP Flybase;
 RG Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN
 RN SEQUENCE FROM N.A.
 RP Flybase;
 RG Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RN
 RN SEQUENCE FROM N.A.
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nunoo J., Pacleb J., Paragas V., Park S., Pounamavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celnikner S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: Contains 5 WD repeats.
 DR EMBL; AE003839; AAM68880.1; -;
 DR EMBL; AY060480; AAL25519.1; -;
 DR InterPro: IPR001680; WD40.
 DR Pfam: PF00400; WD40; 5.
 DR PRINTS; PR00320; GPROTEINBRPT.
 DR ProDom; PD000018; WD40; 2.
 DR SMART; SM00320; WD40; 5.
 DR PROSITE; PS00082; WD_REPEATS_2; 2.
 DR PROSITE; PS00294; WD_REPEATS_REGION; 2.
 KW Repeat; WD repeat.
 SQ
 SEQUENCE 354 AA; 39514 MW; 235785K7B9246C81 CRC64;

Search completed: March 2, 2005, 12:44:08
 Job time : 32.7236 secs

Query Match	Score 41; DB 2; Length 354;
Best Local Similarity	71.4%; Pred. No. 75;
Matches	5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY	1 WXPQFG 7
DB	65 WAHPFG 71

Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 36.0325 Seconds
(without alignments)
85.869 Million cell updates/sec

Title: SEQ1
Perfect score: 50
Sequence: 1 wxhpqfpg 8

Scoring table: BLOSUM62X
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_16dec04:*
1: geneseqp19908:*
2: geneseqp19908:*
3: geneseqp20008:*
4: geneseqp20018:*
5: geneseqp20028:*
6: geneseqp20038:*
7: geneseqp20038:*
8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	50	100.0	8 ABP60360	Abp60360 Streptavi
2	50	100.0	9 AAW32485	Aaw32485 Strept-tag
3	50	100.0	9 AAW59211	Aaw59211 Streptavi
4	50	100.0	9 AAW44010	Aaw44010 Strept pep
5	50	100.0	9 AAY06914	Aay06914 Strept-tag
6	50	100.0	9 AAB30795	Aab30795 Amino aci
7	50	100.0	9 AAG31053	Abg31053 Peptide p
8	50	100.0	9 AAG78476	Aag78476 Strept-tag
9	50	100.0	9 AAO16086	Aao16086 Neurologi
10	50	100.0	9 ABP55547	Abp55547 Strepttag
11	50	100.0	9 AAE37229	Aae37229 Strept-tag
12	50	100.0	9 AAE32860	Aae32860 Strepttag
13	50	100.0	9 AAE33270	Aae33270 Strept epi
14	50	100.0	9 AAG72479	Abg72479 Strept tag
15	50	100.0	9 ADB84587	Adb84587 Streptavi
16	50	100.0	9 ADJ34695	Adj34695 Strept pep
17	50	100.0	9 ADJ36219	Adj36219 Self-coal
18	50	100.0	9 ADJ88138	Adj88138 Strept tag
19	50	100.0	9 ADN29604	Adn29604 Strept tag
20	50	100.0	9 ADQ82347	Adq82347 Strept tag
21	50	100.0	10 AAR66446	Aar66446 Streptavi
22	50	100.0	10 AAW50025	Aaw50025 Peptide t
23	50	100.0	10 AAY57325	Aay57325 Human IL-
24	50	100.0	10 AAY67424	Aay67424 Streptavi
25	50	100.0	10 AAB19071	Aab19071 Peptide b

26	50	100.0	10 4 AAB61168	Aab61168 Paramagne
27	50	100.0	10 4 AAB99026	Aab99026 Streptavi
28	50	100.0	10 4 AAU02084	Aau02084 Streptavi
29	50	100.0	10 5 AAO14759	Aao14759 Human int
30	50	100.0	10 7 ADD29929	Add29929 Antibody-
31	50	100.0	11 2 AAR52691	Aar52691 PASK46-P1
32	50	100.0	11 2 AAR52692	Aar52692 PASK46-P1
33	50	100.0	12 2 AAR52696	Aar52696 PASK46-P1
34	50	100.0	12 2 AAR81821	Aar81821 Streptavi
35	50	100.0	18 5 AAG78473	Aag78473 Linker co
36	50	100.0	20 6 ADA37002	Ada37002 Artificial
37	50	100.0	23 5 AAO19984	Aao19984 Peptide o
38	50	100.0	42 5 AAO19988	Aao19988 Protein o
39	50	100.0	42 5 AAO19986	Aao19986 Protein o
40	50	100.0	46 8 ADJ32868	Adj32868 Receptor
41	50	100.0	46 8 ADJ55658	Adj55658 ED Peptid
42	50	100.0	47 2 AAR51892	Aar51892 VHI leade
43	50	100.0	47 5 ABB81776	Abb81776 Plasmid p
44	50	100.0	132 8 ADJ36263	Adj36263 Self-coal
45	50	100.0	155 2 AAW32483	Aaw32483 Kappa 11g

ALIGNMENTS

RESULT 1
ID ABP60360 standard; peptide; 8 AA.
XX
AC ABP60360;
XX
DT 28-MAR-2003 (first entry)
XX
DE Streptavidin tag peptide SEQ ID NO 1.
XX
KM Streptavidin; protein chip; microtitre plate; detection.
XX
OS Synthetic.
XX
PN DE10113776-A1.
XX
PD 02-OCT-2002.
XX
PP 21-MAR-2001; 2001DE-01013776.
XX
PR 21-MAR-2001; 2001DE-01013776.
XX
PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
PI Schmidt T;
XX
DR WPI; 2003-031166/03.
XX
PT New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.
XX
PS Disclosure; Page 2; 18pp; German.
XX
CC The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it may not be essential to remove it); facilitates detection and has easily controllable binding properties. (I) is particularly used for purifying FP from dilute solution in batch formats (which use simpler apparatus than column methods and result in lower loss of FP). The present sequence is that of a streptavidin tag peptide disclosed with the invention

XX
SQ Sequence 8 AA;

Query Match 100.0%; Score 50; DB 6; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.8e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WKHPQFG 8
DB 1 WKHPQFG 8

RESULT 2
AAW32485
ID AAW32485 standard; peptide; 9 AA.

XX AC AAW32485;

XX DT 27-MAR-1998 (first entry)

XX DE Strep-tag peptide.

XX KM Catalytic antibody; growth factor; B-cell mitogenesis; LHL; L protein;
XX KW hen egg lysozyme; strep-tag; purification.

XX OS Synthetic.

XX PN WO9735887-A1.

XX PD 02-OCT-1997.

XX PF 26-MAR-1997; 97WO-AU000194.

XX PR 26-MAR-1996; 96AU-00008951.

XX PR 27-FEB-1997; 97AU-00005375.

XX PA (AMRA-) AMRAD OPERATIONS PTY LTD.

XX PI Koentgen F, Suesse GM, Tarlinton DM, Treutlein HR;

XX DR WPI; 1997-489572/45.

XX PT New catalytic antibody precursors - comprising a B-cell surface molecule
XX PT binding portion which can induce B-cell mitogenesis.

XX PS Example 3; Page 40; 109pp; English.

XX CC This peptide comprises the strep-tag peptide that is recognised by
XX CC streptavidin. A form of novel growth factor LHL (see AAW32479) was
XX CC generated by PCR that contains a FLAG epitope (see AAW32484) at its N-
XX CC terminus and the strep-tag at its C-terminus. The construct is designated
XX CC LHL-seq (see AAW32481), where L is the immunoglobulin binding entity from
XX CC Reprosterococcus magnus and H is residues 42-62 of hen egg lysozyme.
XX CC The strep-tag was used for purification of LHL-seq over a streptavidin
XX CC column. Thus, the LHL-seq was not purified on the basis of binding
XX CC immunoglobulin, thereby eliminating potential contamination by other
XX CC bacterial proteins which also bind immunoglobulin. LHL-seq has identical
XX CC activity to that of LHL, and can be used in novel methods for the
XX CC generation of catalytic antibodies

XX SQ Sequence 9 AA;

Query Match 100.0%; Score 50; DB 2; Length 9;
Best Local Similarity 87.5%; Pred. No. 1.8e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WKHPQFG 8
DB 2 WKHPQFG 9

RESULT 3
AAW59211

ID AAW59211 standard; peptide; 9 AA.

XX AC AAW59211;

XX DT 27-AUG-1998 (first entry)

XX DE Streptavidin tagged peptide ligand #1.

XX KM Streptavidin; ligand; binding affinity; mutant; isolation; purification;
XX KW recover; immobilise.

XX OS Synthetic.

XX PN EP835934-A2.

XX PD 15-APR-1998.

XX PF 09-OCT-1997; 97EP-00117504.

XX PR 10-OCT-1996; 96DE-01041876.

XX PA (BIOA-) INST BIOANALYTIK GMBH.

XX PI Skerra A, Voss S;

XX DR WPI; 1998-218868/20.

XX PT Streptavidin mutants with higher binding affinity for peptide ligands -
XX PT have mutation in amino acid region 44-53, used to isolate, purify or
XX PT determine fusion proteins including these ligands.

XX PS Claim 10; Page 11; 21pp; German.

XX CC AAW59211 and AAW59212 are ligands used in a method to assay binding
XX CC affinity of streptavidin mutants. These mutants have a mutation within
XX CC the amino acid (aa) region 44-53 of the wild-type protein show a higher
XX CC binding affinity than the wild-type for peptide ligands that include the
XX CC sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z
XX CC are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin
XX CC mutants can be used to isolate, purify and determine proteins or to
XX CC determine/recover substances that contain streptavidin-binding groups.
XX CC Such compounds may also be used to immobilise fusions on microtitre
XX CC plates, microbeads or sensor chips

XX SQ Sequence 9 AA;

Query Match 100.0%; Score 50; DB 2; Length 9;
Best Local Similarity 87.5%; Pred. No. 1.8e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WKHPQFG 8
DB 2 WKHPQFG 9

RESULT 4
AAW44010
ID AAW44010 standard; peptide; 9 AA.

XX AC AAW44010;

XX DT 05-JUN-1998 (first entry)

XX DE Strep peptide epitope used in an epitope tagged prion protein construct.

XX KM Prion; epitope; FLAG; Strep; poly-histidine; haemagglutinin; recombinant;
XX KW transgenic animal; scrapie; Creutzfeldt-Jakob disease; CJD;
XX KW bovine spongiform encephalopathy; BSE.

XX OS Synthetic.

XX PN WO9746572-A1.

```

PD 11-DEC-1997;
XX
XX 29-MAY-1997; 97WO-US009289.
XX
XX 06-JUN-1996; 96US-00660626.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Prusiner SB, Telling GC, Cohen FE, Scott MR;
XX
XX WPI; 1998-042112/04.
XX
XX Nucleic acid construct encoding biologically active protein and epitope -
XX especially epitope-tagged prion protein.
XX
XX Claim 4; Page 48; 62pp; English.
XX
XX This sequence represents an artificial Strep peptide epitope. It is used
XX in a recombinant nucleic acid construct encoding an epitope-tagged prion
XX protein (PrP). The construct comprises a first nucleic acid sequence
XX encoding an amino acid sequence of a biologically active protein fragment
XX and a second nucleic acid sequence encoding a heterologous epitope
XX domain. The heterologous epitope domain is a peptide selected from a
XX peptide group of FlAg, Strep, poly-histidine, human c-myc peptide
XX recognised by monoclonal antibody 9E10 and haemagglutinin peptide
XX recognised by monoclonal antibody 12CA5. The protein is a natural,
XX synthetic or chimeric PrP molecule. The protein has two different three-
XX dimensional conformations and the epitope domain is spatially positioned
XX relative to the protein such that the epitope domain is more exposed in a
XX first conformation relative to a second conformation. The nucleic acid
XX construct may be used for the production of transgenic animals or cells
XX that are useful in a method for distinguishing between different
XX conformational shapes of a protein. These methods are particularly useful
XX in studying diseases caused by prion proteins, e.g. Creutzfeldt-Jakob
XX disease (CJD), scrapie and bovine spongiform encephalopathy (BSE)
XX
XX Sequence 9 AA;
XX
XX Query Match 100.0%; Score 50; DB 2; Length 9;
XX Best Local Similarity 87.5%; Pred. No. 1.8e+06;
XX Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 WKHPQFGG 8
XX |:|||||
XX Db 2 WRHPQFGG 9
XX
XX RESULT 5
XX ID AAY06914 standard; peptide; 9 AA.
XX
XX AAY06914;
XX
XX 01-JUL-1999 (first entry)
XX
XX Strep-tag sequence.
XX
XX Growth factor precursor; B-cell surface; T cell surface; CAB; hepatitis;
XX catalytic antibody; immunoglobulin; tumour necrosis factor; influenza;
XX rheumatoid arthritis; toxic shock syndrome; viral docking receptor;
XX human immune virus; tumour-specific antigen; amyloid plaque; myeloma;
XX Alzheimer's disease; IGE; allergy; asthma; drug detoxification;
XX autoimmune; inflammatory disease; gene therapy; strep-tag.
XX
XX Synthetic.
XX
XX WO9915563-A1.
XX
XX 01-APR-1999.
XX
XX 18-SEP-1998; 98WO-AU000783.
XX
XX 19-SEP-1997; 97AU-00009306.
XX

```

```

XX
XX (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
XX Koentgen F, Suesse GM, Tarlinton DM, Treutlein HR;
XX
XX WPI; 1999-244394/20.
XX
XX Growth factor precursor cleaved by antigen-specific catalytic antibody.
XX
XX Example 3; Page 42; 101pp; English.
XX
XX The invention relates to a growth factor precursor that comprises B-cell
XX surface binding part, T cell surface binding part, antigen cleavable by a
XX catalytic antibody (CAB); and a peptide comprising heavy and light chains
XX of immunoglobulin. When the antigen is cleaved the B cell surface part
XX can interact with its target. The growth factor precursors are used to
XX select B cells that produce Ag-specific CAB, and to generate CAB from
XX such cells (by inducing mitogenesis, caused by the growth factor released
XX by specific cleavage). The Ag-specific CAB can be directed against, e.g.
XX tumour necrosis factor (treatment of rheumatoid arthritis or toxic shock
XX syndrome); viral docking receptors (treatment of human immune virus,
XX hepatitis and influenza infections); tumour-specific antigens; amyloid
XX plaque (treatment of Alzheimer's disease or myeloma) or IGE (treatment of
XX allergies such as asthma). CAB may also be used for drug detoxification,
XX to treat autoimmune or inflammatory diseases and to eliminate
XX environmental or industrial pollutants, such as plastics and petroleum.
XX Particularly the growth factor precursors are produced by delivering the
XX corresponding nucleic acid in a viral or other gene therapy vector. The
XX present sequence represents a strep-tag
XX
XX Sequence 9 AA;
XX
XX Query Match 100.0%; Score 50; DB 2; Length 9;
XX Best Local Similarity 87.5%; Pred. No. 1.8e+06;
XX Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 WKHPQFGG 8
XX |:|||||
XX Db 2 WRHPQFGG 9
XX
XX RESULT 6
XX ID AAB30795 standard; peptide; 9 AA.
XX
XX AAB30795;
XX
XX 02-APR-2001 (first entry)
XX
XX Amino acid sequence of a Strep epitope.
XX
XX SCHAG; self-coalesce; higher-order aggregate; amyloidogenic domain;
XX aggregation; fibril; phenotypic alteration; gene therapy;
XX disease resistance; plant pigmentation; prion disease.
XX
XX Synthetic.
XX
XX WO200075324-A2.
XX
XX 14-DEC-2000.
XX
XX 09-JUN-2000; 2000WO-US015876.
XX
XX 09-JUN-1999; 99US-0138833P.
XX
XX (ARCH-) ARCH DEV CORP.
XX
XX Lindquist S, Li L, Ma J, Liu J, Sondheimer N, Scheibel T;
XX
XX WPI; 2001-061723/07.
XX
XX New nucleic acid encoding chimeric proteins with self-assembly
XX properties, useful e.g. for diagnosis and treatment of prion diseases,
XX

```


Query Match 100.0%; Score 50; DB 5; Length 9;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFG 8
 |:|||||
 DB 2 WRHPQFG 9

RESULT 9
 AA016086
 ID AA016086 standard; peptide; 9 AA.
 XX
 AC AA016086;
 XX

DT 27-FEB-2003 (first entry)
 XX
 DE Neurological/CNS disease treatment method-related peptide #18.
 XX

KW Vaccine; gene therapy; neurological disease; CNS disorder;
 KM central nervous system disorder; olfactory system; Alzheimer's disease;
 KW Creutzfeld-Jakob disease; Huntington's chorea; Parkinson's disease;
 KM viral infection of the brain; brain tumour; lysosomal storage disease;
 XX multiple sclerosis.
 XX

OS Unidentified.

PN WO200274243-A2.

PD 26-SEP-2002.

PF 15-MAR-2002; 2002WO-US008042.

PR 15-MAR-2001; 2001US-00808037.

PS (UYBA-) UNIT RAMOT APPLIED RES & IND DEV LTD.
 PA (MCIN/) MCINNIS P.

PI Solomon B, Frenkel D;

DR WPI; 2003-040542/03.

PT Treating or diagnosing neurological diseases of the central nervous
 PT system, e.g. Alzheimer's disease, comprises displaying a polypeptide or
 PT diagnostic agent on viral display vehicle and introducing or detecting
 PT the display vehicle.
 PT

PS Disclosure; Page 214; 214pp; English.

CC The invention comprises a method for treating a neurological disease or a
 CC central nervous system (CNS) disorder. The method involves displaying a
 CC therapeutic molecule capable of treating the neurological disease or CNS
 CC disorder on a viral display vehicle. The viral display vehicle is then
 CC introduced into the olfactory system of a subject to treat the disease or
 CC disorder. The method of the invention is useful for preventing, treating
 CC and diagnosing neurological diseases or CNS disorders, such as:
 CC Alzheimer's disease; Creutzfeld-Jakob disease; Huntington's chorea; viral
 CC infections of the brain; brain tumours; lysosomal storage diseases;
 CC Parkinson's disease; and multiple sclerosis. The present amino acid
 CC sequence represents a peptide which was used in the invention
 XX

SO Sequence 9 AA;

Query Match 100.0%; Score 50; DB 6; Length 9;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFG 8
 |:|||||
 DB 2 WRHPQFG 9

RESULT 10

ABP55547
 ID ABP55547 standard; peptide; 9 AA.
 XX

AC ABP55547;

DT 19-FEB-2003 (first entry)
 XX

DE Streptag peptide SEQ ID NO:77.

KW Hepatitis C virus; HCV; envelope protein; hepatotropic; virucide;
 KM antiinflammatory; gene therapy; vaccine.
 XX

OS Hepatitis C virus.
 OS Synthetic.
 OS

PN WO200285932-A2.

PD 31-OCT-2002.

PF 24-APR-2002; 2002WO-BE000062.

PR 24-APR-2001; 2001EP-00870088.

PR 17-JUL-2001; 2001US-0305604P.

PS (INNO-) INNOGENETICS NV.

PI Sablon E, Van Broekhoven A, Bosman A, Depla E, Deschamps G;

DR WPI; 2003-093095/08.

PT New recombinant nucleic acids for expressing Hepatitis C virus (HCV)
 PT envelope proteins in eukaryotic cells, comprising a sequence encoding a
 PT protein having an avian lysozyme leader peptide joined to the HCV
 PT envelope protein.
 PT

PS Claim 5; Page 303; 319pp; English.

CC The present invention describes a recombinant nucleic acid (1) comprising
 CC a nucleotide sequence encoding a protein having an avian lysozyme leader
 CC peptide, or its functional equivalent, joined to a Hepatitis C virus
 CC (HCV) envelope protein or its part. Also described: (1) a vector
 CC comprising the recombinant nucleic acid; (2) a host cell comprising the
 CC recombinant nucleic acid or the vector; and (3) a method for producing
 CC HCV envelope protein or its part in a host cell, comprising transforming
 CC the host cell with the recombinant nucleic acid or with the vector, where
 CC the host cell is capable of expressing the protein cited above. (1) has
 CC hepatotropic, virucide and antiinflammatory activities, and can be used
 CC in gene therapy and vaccines. The recombinant nucleic acid is useful for
 CC efficient expression of Hepatitis C virus envelope proteins in eukaryotic
 CC cells, such as yeast cells. The HCV envelope proteins may be used as a
 CC vaccine, or for incorporation into an immunoassay for the detection of
 CC anti-HCV antibodies, and/or genotyping of HCV, for prognosing or
 CC monitoring of HCV disease, or as a therapeutic agent. The method is used
 CC for producing HCV envelope protein or its part in a host cell. AB084197
 CC to AB084253 and ABP55528 to ABP55568 represent sequences used in the
 CC exemplification of the present invention
 XX

SO Sequence 9 AA;

Query Match 100.0%; Score 50; DB 6; Length 9;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFG 8
 |:|||||
 DB 2 WRHPQFG 9

RESULT 11
 AA037229
 ID AA037229 standard; peptide; 9 AA.
 XX

AC AA037229;

XX 07-AUG-2003 (first entry)
 XX
 DT Strep-tag peptide.
 DE
 XX Gene expression; therapy; isolation.
 KM
 XX Synthetic.
 OS
 XX WO2003038049-A2.
 PN
 XX 08-MAY-2003.
 PD
 XX 29-OCT-2002; 2002WO-US034645.
 PF
 XX 29-OCT-2001; 2001US-0340689P.
 PR
 XX (RENO-) RENOVIS INC.
 PA
 XX Heintz N, Serafini TA, Shyjan AW;
 PI
 XX WPI; 2003-430512/40.
 DR
 XX
 PT Isolating cell-type specific mRNAs, useful in gene expression analysis or
 PT quantification in a specific cell in a heterogeneous cell mixture, by
 PT isolating mRNA bound to ribosomes or proteins that bind mRNA in a cell
 PT type specific manner.
 CC
 XX Disclosure; Page 8; 136pp; English.
 PS
 XX The invention relates to a method for isolating mRNA from a population of
 CC cells. The method involves selectively isolating ribosomes or proteins
 CC that bind mRNA in a cell type specific manner and then isolating the mRNA
 CC bound to the ribosomes or proteins that bind mRNA. The method is useful
 CC for facilitating the analysis and quantification of gene expression in a
 CC selected cell type present within a heterogeneous cell mixture. The
 CC method may also be used in diagnostics or therapies for human diseases.
 CC The present sequence is Strep-tag peptide. This sequence is used to
 CC illustrate the method of the invention
 CC
 SQ Sequence 9 AA;
 XX
 XX
 Query Match 100.0%; Score 50; DB 6; Length 9;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WXHRPQFG 8
 DB 2 WXHRPQFG 9

PR 17-JUL-2001; 2001US-0305604P.
 XX
 XX (INNO-) INNOGENETICS NV.
 PA
 XX Depia E, Bosman A, Deschamps G, Sablon E, Suckow M, Samson I,
 PI Verheyden G;
 XX WPI; 2003-103409/09.
 DR
 XX
 PT New Hepatitis C virus (HCV) envelope protein with N-glycosylation
 PT site(s), useful as a vaccine for inducing a HCV-specific immune response
 PT or HCV-specific antibodies, particularly for preventing or treating HCV
 PT infection.
 CC
 XX Claim 10; Page 336; 355pp; English.
 PS
 XX The invention relates to Hepatitis C virus (HCV) envelope protein with N-
 CC glycosylation site(s). The HCV envelope protein, or the pharmaceutical
 CC composition comprising the envelope protein, is useful as a medicament or
 CC a vaccine, particularly for inducing a HCV-specific immune response,
 CC including HCV-specific antibodies or inducing a T-cell function in a
 CC mammal. The protein is particularly useful for preventing, treating or
 CC diagnosing HCV infection. It is also useful for detecting the presence of
 CC anti-HCV antibodies in a sample. The present sequence is a peptide used
 CC in the invention
 CC
 SQ Sequence 9 AA;
 XX
 XX
 Query Match 100.0%; Score 50; DB 6; Length 9;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WXHRPQFG 8
 DB 2 WXHRPQFG 9

RESULT 13
 ID AAE33270 standard; peptide; 9 AA.
 XX
 AC AAE33270;
 XX
 DT 02-APR-2003 (first entry)
 DE
 XX Strep epitope tag used to illustrate the method of the invention.
 XX
 KM Membrane-spanning signal-transducing protein; MSST protein; epitope.
 KW
 XX Synthetic.
 OS
 XX WO200286507-A1.
 PN
 XX 31-OCT-2002.
 PD
 XX 24-APR-2002; 2002WO-US013250.
 PF
 XX 24-APR-2001; 2001US-0286350P.
 PR
 XX 21-AUG-2001; 2001US-00935061.
 PA
 XX (STRD) UNIV IELAND STANFORD JUNIOR.
 PI
 XX Kobilka BK, Ghanouni P, Lee TW;
 XX WPI; 2003-103418/09.
 DR
 XX
 PT Identifying an agent that modulates activity of a membrane-spanning,
 PT signal-transducing (MSST) protein, by detecting a conformational change
 PT in a MSST protein upon interaction with a ligand.
 CC
 PS Disclosure; Page 84; 104pp; English.
 XX
 XX The present invention relates to methods and compositions for identifying

CC Agents that modulate activity of a membrane-spanning, signal-transducing
 CC (MST) protein. The method involves detecting a conformational change in
 CC a MST protein upon interaction with a ligand. The method is useful for
 CC identifying agents that modulate (e.g. agonists or antagonists) activity
 CC of MST protein. The present sequence is a strep epitope tag used to
 CC illustrate the method of the invention

XX Sequence 9 AA:

Query Match 100.0%; Score 50; DB 6; Length 9;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXPQFG 8
 :|:|:|:|
 Db 2 WRHPQFG 9

RESULT 14

ID ABG72479 standard; peptide; 9 AA.

XX ABG72479;

DT 18-FEB-2003 (first entry)

XX Strep tag for use as a G protein epitope tag.

DE G protein; alpha sub-unit; G protein coupled receptor; GPCR;
 KW G protein activation; G protein coupled receptor activation;
 KM G protein coupled receptor identification; strep tag.

XX Synthetic.

XX US6448377-B1.

XX 10-SEP-2002.

XX 27-SEP-2000; 2000US-00672239.

XX 27-SEP-2000; 2000US-00672239.

XX (STRD) UNIV. IELAND STANFORD JUNIOR.

XX Kobilka B, Lee TW;

XX WPI; 2003-110149/10.

PT Determining effects of candidate agent on activation of a G protein
 PT coupled receptor (GPCR) for evaluating new agonists and/or inverse
 PT agonists for GPCR by contacting a candidate agent with a modified G
 PT protein alpha subunit and a GPCR.

XX Disclosure; Col 9; 38pp; English.

CC The invention describes a method of determining the effects of a
 CC candidate agent on activation of a G protein coupled receptor (GPCR). The
 CC method comprises contacting a candidate agent with a modified G protein
 CC alpha subunit and a GPCR, and detecting a level of G protein activation
 CC in response to the contacting, where the level of activation is
 CC indicative of the effects of the agent on the activity of GPCR. The
 CC method is useful for determining the effects of a candidate agent on
 CC activation of a G protein coupled receptor, evaluating new agonists,
 CC and/or inverse agonists for GPCRs, identifying ligands for GPCRs, and
 CC developing a strategy for identifying GPCRs involved in different
 CC biological processes, including diseases. The invention provides rapid
 CC and more sensitive bioassays for evaluating new agonists, agonists and/or
 CC inverse agonists for GPCR. The method can be performed using membranes,
 CC which increases both the ease of performing the assay and its efficacy,
 CC and also allows high throughput screening of GPCR activity. Furthermore,
 CC this method directly measures GPCR activity, and thus is less labour-
 CC intensive than the conventional methods. This sequence represents the
 CC Strep tag, an example of an epitope tag that can be used to tether the

CC stimulatory G protein alpha sub-unit to a membrane

XX Sequence 9 AA:

Query Match 100.0%; Score 50; DB 6; Length 9;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXPQFG 8
 :|:|:|:|
 Db 2 WRHPQFG 9

RESULT 15

ID ADB84587 standard; peptide; 9 AA.

XX ADB84587;

DT 04-DEC-2003 (first entry)

XX Streptavidin conserved peptide #1.

DE cell-free transcription system; cell-free translation system;
 KW protein synthesis; matrix; streptavidin.

XX Escherichia coli.

XX DE10137792-A1.

XX 27-FEB-2003.

XX 06-AUG-2001; 2001DE-01037792.

XX 06-AUG-2001; 2001DE-01037792.

XX (ERDM/) ERDMANN V.

XX Erdmann VA, Lamla T, Stiege W;

XX WPI; 2003-343999/33.

PT Expressing genes in cell-free system, useful for preparation of proteins,
 PT comprises that the protein formed is removed from solution by binding to
 PT a matrix.

XX Claim 13; Col 8; 8pp; German.

CC This invention describes a novel method of expressing genes in a cell-
 CC free transcription and translation system which comprises using a
 CC reaction solution containing all necessary components of the
 CC transcription/translation system, amino acids, nucleotides and
 CC metabolites that supply energy and that are needed for synthesis. The
 CC proteins formed are immobilised on a matrix. The method allows simple
 CC recovery of proteins without a separate isolation step and the amount of
 CC proteins produced can be determined before a reaction is complete.

CC Continuous removal of proteins prevents it interfering with the
 CC expression process, making possible synthesis of proteins that interact
 CC adversely with the process, so normally produced only in very low yields.
 CC The use of a matrix concentrates the proteins produced and is applicable
 CC to proteins of any size. This sequence represents a highly conserved
 CC peptide from E. coli streptavidin which is used to illustrate the method
 CC of the invention.

XX Sequence 9 AA:

Query Match 100.0%; Score 50; DB 7; Length 9;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXPQFG 8
 :|:|:|:|
 Db 2 WRHPQFG 9

Wed Mar 2 14:27:02 2005

seq1.rag

Page 8

Search completed: March 2, 2005, 13:02:42
Job time : 39.0325 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 24.878 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ1
Perfect score: 50
Sequence: 1 wxhpgfgg 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues
Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	8	US-10-026-578B-1	Sequence 1, Appli
2	50	100.0	9	US-09-808-037-32	Sequence 32, Appli
3	50	100.0	9	US-09-983-067-2	Sequence 2, Appli
4	50	100.0	9	US-09-935-061-3	Sequence 3, Appli
5	50	100.0	9	US-10-128-590-77	Sequence 77, Appli
6	50	100.0	9	US-10-345-618-14	Sequence 14, Appli
7	50	100.0	9	US-10-128-587A-77	Sequence 77, Appli
8	50	100.0	9	US-10-384-788-32	Sequence 32, Appli
9	50	100.0	9	US-10-449-831A-166	Sequence 166, App
10	50	100.0	9	US-10-425-000-75	Sequence 24, Appli
11	50	100.0	9	US-10-424-999-24	Sequence 24, Appli
12	50	100.0	9	US-10-272-196-30	Sequence 30, Appli
13	50	100.0	9	US-10-612-410-18	Sequence 18, Appli

14	50	100.0	9	US-10-692-071-3	Sequence 3, Appli
15	50	100.0	9	US-10-397-438A-6	Sequence 6, Appli
16	50	100.0	9	US-10-753-309-5	Sequence 5, Appli
17	50	100.0	9	US-10-494-248-5	Sequence 5, Appli
18	50	100.0	10	US-09-801-968-18	Sequence 18, Appli
19	50	100.0	10	US-09-802-154-18	Sequence 18, Appli
20	50	100.0	10	US-10-060-765-13	Sequence 13, Appli
21	50	100.0	10	US-10-263-230A-12	Sequence 12, Appli
22	50	100.0	10	US-10-818-140-13	Sequence 13, Appli
23	50	100.0	10	US-10-771-173-13	Sequence 9, Appli
24	50	100.0	12	US-09-904-599A-8	Sequence 8, Appli
25	50	100.0	18	US-10-397-438A-9	Sequence 9, Appli
26	50	100.0	20	US-10-342-805-1	Sequence 1, Appli
27	50	100.0	23	US-10-344-607-2	Sequence 2, Appli
28	50	100.0	42	US-10-344-607-16	Sequence 16, Appli
29	50	100.0	42	US-10-344-607-20	Sequence 20, Appli
30	50	100.0	46	US-10-422-262-12	Sequence 12, Appli
31	50	100.0	46	US-10-448-609-8	Sequence 8, Appli
32	50	100.0	47	US-10-416-290-57	Sequence 57, Appli
33	50	100.0	132	US-10-449-831A-212	Sequence 212, App
34	50	100.0	135	US-10-345-618-11	Sequence 11, Appli
35	50	100.0	178	US-10-345-618-13	Sequence 13, Appli
36	50	100.0	194	US-10-416-708A-10	Sequence 10, Appli
37	50	100.0	194	US-10-416-708A-64	Sequence 64, Appli
38	50	100.0	198	US-10-345-618-8	Sequence 8, Appli
39	50	100.0	210	US-09-272-809-24	Sequence 24, Appli
40	50	100.0	212	US-10-449-831A-198	Sequence 198, App
41	50	100.0	234	US-10-449-831A-192	Sequence 192, App
42	50	100.0	329	US-10-449-831A-204	Sequence 204, App
43	50	100.0	334	US-10-017-736-4	Sequence 4, Appli
44	50	100.0	334	US-10-650-585-4	Sequence 4, Appli
45	50	100.0	342	US-10-345-618-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-10-026-578B-1
; Sequence 1, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-1

Query Match 100.0%; Score 50; DB 14; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.3e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 wxhpgfgg 8
DB 1 wxhpgfgg 8

RESULT 2
US-09-808-037-32
; Sequence 32, Application US/09808037
; Patent No. US20020052311A1
; GENERAL INFORMATION:
; APPLICANT: SOLOMON, Beke
; APPLICANT: HANAN, Eliat
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE TREATMENT AND/OR DIAGNOSIS OF
; FILE REFERENCE: NEUROLOGICAL DISEASES AND DISORDERS
; CURRENT APPLICATION NUMBER: US/09/808,037
; PRIOR FILING DATE: 2001-03-15
; PRIOR FILING DATE: 09/629,971
; PRIOR FILING DATE: 2000-07-31
; PRIOR FILING DATE: 1999-12-29
; PRIOR FILING DATE: 1999-09-03
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 32
; LENGTH: 9
; TYPE: PRT
; ORGANISM: synthetic construct
US-09-808-037-32

Query Match 100.0%; Score 50; DB 9; Length 9;
Best Local Similarity 87.5%; Pred. No. 1.3e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPPOFG 8
|:|||||
Db 2 WRHPQFG 9

RESULT 3
US-09-983-067-2
; Sequence 2, Application US/09983067
; Patent No. US20020123101A1
; GENERAL INFORMATION:
; APPLICANT: INOUE, Akio
; APPLICANT: SHIMIZU, Yoshihiro
; APPLICANT: USDA, Takuya
; TITLE OF INVENTION: Process for Producing Peptides By Using In Vitro
; FILE REFERENCE: 1752-0151P
; CURRENT APPLICATION NUMBER: US/09/983,067
; PRIOR FILING DATE: 2001-10-23
; PRIOR FILING DATE: 2001-09-26
; PRIOR FILING DATE: 2001-07-27
; PRIOR FILING DATE: 2001-07-27
; PRIOR FILING DATE: 2001-01-15
; PRIOR FILING DATE: 2001-01-15
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Strept-tag binding to streptavidin
; NAME/KEY: MISC FEATURE
; LOCATION: (1)..(9)
; OTHER INFORMATION: Schmidt & Skerra, 1993, "The random peptide library-assisted
; OTHER INFORMATION: engineering of a C-terminal affinity peptide, useful for the
; OTHER INFORMATION: detection and purification of a functional IgFv fragment",
US-09-983-067-2

Query Match 100.0%; Score 50; DB 9; Length 9;
Best Local Similarity 87.5%; Pred. No. 1.3e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPPOFG 8
|:|||||
Db 2 WRHPQFG 9

RESULT 4
US-09-935-061-3
; Sequence 3, Application US/09935061
; Publication No. US20030129649A1
; GENERAL INFORMATION:
; APPLICANT: Kodlika, Brian M.
; APPLICANT: Ghannouni, Bejman
; APPLICANT: Lee, Tae Weon
; TITLE OF INVENTION: Conformational assays to detect binding
; FILE REFERENCE: to G protein-coupled receptors
; CURRENT APPLICATION NUMBER: US/09/935,061
; PRIOR FILING DATE: 2001-08-21
; PRIOR FILING DATE: 2001-04-24
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: epitope tag peptide
US-09-935-061-3

Query Match 100.0%; Score 50; DB 10; Length 9;
Best Local Similarity 87.5%; Pred. No. 1.3e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPPOFG 8
|:|||||
Db 2 WRHPQFG 9

RESULT 5
US-10-128-590-77
; Sequence 77, Application US/10128590
; Publication No. US20030108561A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins
; FILE REFERENCE: 135 PCT
; CURRENT APPLICATION NUMBER: US/10/128,590
; PRIOR FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77
; LENGTH: 9
; TYPE: PRT
; ORGANISM: streptag
US-10-128-590-77

Query Match 100.0%; Score 50; DB 14; Length 9;
Best Local Similarity 87.5%; Pred. No. 1.3e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPPOFG 8
|:|||||
Db 2 WRHPQFG 9

RESULT 6
US-10-345-618-14
; Sequence 14, Application US/10345618

```
; Publication No. US2003014848A1
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Sues, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF PRODUCING SAME
; FILE REFERENCE: 13474
; CURRENT APPLICATION NUMBER: US/10/345,618
; CURRENT FILING DATE: 2003-01-16
; PRIOR APPLICATION NUMBER: US/09/509,031
; PRIOR FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Strep-tag
US-10-345-618-14

Query Match          100.0%; Score 50; DB 14; Length 9;
Best Local Similarity 87.5%; Pred.No. 1.3e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHRPQFG 8
   |||||
   |||||
Db 2 WXHRPQFG 9

RESULT 7
US-10-128-587A-77
; Sequence 77, Application US/10128587A
; Publication No. US20030152940A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope
; FILE REFERENCE: 134 PCT
; CURRENT APPLICATION NUMBER: US/10/128,587A
; CURRENT FILING DATE: 2002-04-24
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: streptag
US-10-128-587A-77

Query Match          100.0%; Score 50; DB 14; Length 9;
Best Local Similarity 87.5%; Pred.No. 1.3e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHRPQFG 8
   |||||
   |||||
Db 2 WXHRPQFG 9

RESULT 8
US-10-384-788-32
; Sequence 32, Application US/10384788
; Publication No. US20040013647A1
; GENERAL INFORMATION:
; APPLICANT: SOLOMON, Beke
; APPLICANT: PRENKEL, Dan
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING A PLAQUE-FORMING DISEASE
; FILE REFERENCE: SOLOMON=2D.2
; CURRENT APPLICATION NUMBER: US/10/384,788
; CURRENT FILING DATE: 2003-03-11
; PRIOR APPLICATION NUMBER: 60/371,735
```

```
; PRIOR FILING DATE: 2002-04-12
; PRIOR APPLICATION NUMBER: 09/808,037
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 09/830,954
; PRIOR FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 10/162,889
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: 09/473,653
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/629,971
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 60/152,417
; PRIOR FILING DATE: 1999-09-03
; PRIOR APPLICATION NUMBER: PCT/IL00/00518
; PRIOR FILING DATE: 2000-08-31
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 32
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Synthetic construct
US-10-384-788-32

Query Match          100.0%; Score 50; DB 15; Length 9;
Best Local Similarity 87.5%; Pred.No. 1.3e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHRPQFG 8
   |||||
   |||||
Db 2 WXHRPQFG 9

RESULT 9
US-10-449-831A-166
; Sequence 166, Application US/10449831A
; Publication No. US20040029179A1
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; TITLE OF INVENTION: Higher molecular weight entities and uses therefor
; FILE REFERENCE: 2385978
; CURRENT APPLICATION NUMBER: US/10/449,831A
; CURRENT FILING DATE: 2003-05-30
; PRIOR APPLICATION NUMBER: USSN 60/384878
; PRIOR FILING DATE: 2002-05-31
; NUMBER OF SEQ ID NOS: 237
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 166
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Strep tag
US-10-449-831A-166

Query Match          100.0%; Score 50; DB 15; Length 9;
Best Local Similarity 87.5%; Pred.No. 1.3e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHRPQFG 8
   |||||
   |||||
Db 2 WXHRPQFG 9

RESULT 10
US-10-425-000-75
; Sequence 75, Application US/10425000
; Publication No. US20040052777A1
; GENERAL INFORMATION:
; APPLICANT: Nesbitt, Mark
; APPLICANT: Cameron, Beatrice
; APPLICANT: Blanche, Francis
; TITLE OF INVENTION: Kringles Polypeptides and Methods for Using Them to Inhibit
; TITLE OF INVENTION: Angiogenesis
```

FILE REFERENCE: ST01027-B
CURRENT APPLICATION NUMBER: US/10/425,000
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: 10/233,675
PRIOR FILING DATE: 2002-09-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: PatentIn version 3.2
SEQ ID NO 75
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Purification tag
US-10-425-000-75

Query Match
Best Local Similarity 100.0%; Score 50; DB 15; Length 9;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFG 8
|:|||||
Db 2 WRHPQFG 9

RESULT 11
US-10-424-999-24
Sequence 24, Application US/10424999
Publication No. US20040052810A1
GENERAL INFORMATION:
APPLICANT: Nesbitt, Mark
APPLICANT: Cameron, Beatrice
TITLE OF INVENTION: Abrogen Polypeptides, Nucleic Acids Encoding Them and Methods for
FILE REFERENCE: ST01027-A
CURRENT APPLICATION NUMBER: US/10/424,999
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: 10/233,675
PRIOR FILING DATE: 2002-09-04
NUMBER OF SEQ ID NOS: 70
SOFTWARE: PatentIn version 3.2
SEQ ID NO 24
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Purification tag
US-10-424-999-24

Query Match
Best Local Similarity 100.0%; Score 50; DB 15; Length 9;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFG 8
|:|||||
Db 2 WRHPQFG 9

RESULT 12
US-10-272-196-30
Sequence 30, Application US/10272196
Publication No. US2004007246A1
GENERAL INFORMATION:
APPLICANT: Murray, Claire Margaret
APPLICANT: Hutchinson, Raymond
APPLICANT: Bantick, John Raymond
APPLICANT: Sullivan, Michael
APPLICANT: Donald, David Keith
APPLICANT: Jackson, Andrew Paul
APPLICANT: Cook, Ian David
TITLE OF INVENTION: INHIBITORS OF MONOCARBOXYLATE TRANSPORT
FILE REFERENCE: 06275-285001

CURRENT APPLICATION NUMBER: US/10/272,196
CURRENT FILING DATE: 2002-10-16
PRIOR APPLICATION NUMBER: US 60/329,318
PRIOR FILING DATE: 2001-10-16
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PasteSeq for Windows Version 4.0
SEQ ID NO 30
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
US-10-272-196-30

Query Match
Best Local Similarity 100.0%; Score 50; DB 15; Length 9;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFG 8
|:|||||
Db 2 WRHPQFG 9

RESULT 13
US-10-612-410-18
Sequence 18, Application US/10612410
Publication No. US2004013213A1
GENERAL INFORMATION:
APPLICANT: Bennett, Robert P.
TITLE OF INVENTION: Methods and Compositions for the Production, Identification and
FILE REFERENCE: 0942,5510003
CURRENT APPLICATION NUMBER: US/10/612,410
CURRENT FILING DATE: 2003-07-03
PRIOR APPLICATION NUMBER: 60/393,756
PRIOR FILING DATE: 2002-07-08
PRIOR APPLICATION NUMBER: 60/396,627
PRIOR FILING DATE: 2002-07-19
PRIOR APPLICATION NUMBER: 60/417,172
PRIOR FILING DATE: 2002-10-10
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn version 3.2
SEQ ID NO 18
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Strep epitope
US-10-612-410-18

Query Match
Best Local Similarity 100.0%; Score 50; DB 16; Length 9;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFG 8
|:|||||
Db 2 WRHPQFG 9

RESULT 14
US-10-692-071-3
Sequence 3, Application US/10692071
Publication No. US20040157268A1
GENERAL INFORMATION:
APPLICANT: Koblika, Brian K.
APPLICANT: Ghannouni, Pejman
APPLICANT: Lee, Tae Weon
TITLE OF INVENTION: Conformational assays to detect binding
FILE REFERENCE: STAN-213CIP
CURRENT APPLICATION NUMBER: US/10/692,071
CURRENT FILING DATE: 2003-10-22
PRIOR APPLICATION NUMBER: PCT/US02/13250
PRIOR FILING DATE: 2002-04-24
PRIOR APPLICATION NUMBER: 09/935,061

/ PRIOR FILING DATE: 2001-08-21
/ PRIOR APPLICATION NUMBER: 60/286,250
/ PRIOR FILING DATE: 2001-04-24
/ NUMBER OF SEQ ID NOS: 22
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 3
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: epitope tag peptide
US-10-692-071-3

Query Match 100.0%; Score 50; DB 16; Length 9;
Best Local Similarity 87.5%; Pred.No.1.3e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFG 8
|:|||||
DB 2 WRHPQFG 9

RESULT 15
US-10-397-438A-6
/ Sequence 6, Application US/10397438A
/ Publication No. US20040191869A1
/ GENERAL INFORMATION:
/ APPLICANT: Pharmacia & Upjohn AB
/ TITLE OF INVENTION: Fusion vectors
/ FILE REFERENCE: 00126
/ CURRENT APPLICATION NUMBER: US/10/397,438A
/ CURRENT FILING DATE: 2002-11-05
/ NUMBER OF SEQ ID NOS: 14
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 6
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Strep-tag
/ OTHER INFORMATION: linker
US-10-397-438A-6

Query Match 100.0%; Score 50; DB 16; Length 9;
Best Local Similarity 87.5%; Pred.No.1.3e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFG 8
|:|||||
DB 2 WRHPQFG 9

Search completed: March 2, 2005, 14:18:49
Job time : 26.878 secs
!

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 9.10569 Seconds
(without alignments)
65.585 Million cell updates/sec

Title: SEQ1
Perfect score: 50
Sequence: 1 wxhpqfgg 8

Scoring table:
BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/1/1aa/6CTUS_COMB.pep:*
- 6: /cgn2_6/prodata/1/1aa/Backfile1.pep:*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	9	1	US-08-660-626-3
2	50	100.0	9	3	US-08-828-741B-14
3	50	100.0	9	3	US-08-948-097-1
4	50	100.0	9	3	US-09-031-168-3
5	50	100.0	9	3	US-09-160-567-14
6	50	100.0	9	4	US-09-672-239-3
7	50	100.0	9	4	US-09-710-299-14
8	50	100.0	9	4	US-09-509-031-14
9	50	100.0	9	4	US-09-669-516C-3
10	50	100.0	10	1	US-08-294-386C-11
11	50	100.0	10	2	US-08-737-316A-5
12	50	100.0	10	3	US-08-897-020-6
13	50	100.0	10	3	US-08-895-707-11
14	50	100.0	10	3	US-09-350-823-6
15	50	100.0	10	4	US-09-715-805-13
16	50	100.0	10	4	US-09-692-945-9
17	50	100.0	10	4	US-09-640-041-7
18	50	100.0	10	5	PCT-US95-10224-11
19	50	100.0	12	3	US-08-218-369-8
20	50	100.0	12	4	US-09-504-599A-8
21	50	100.0	12	5	PCT-US95-03742-8
22	50	100.0	15	1	US-08-664-449-33
23	50	100.0	155	3	US-08-828-741B-11
24	50	100.0	155	3	US-09-160-567-11
25	50	100.0	155	4	US-09-710-299-11
26	50	100.0	155	4	US-09-509-031-11
27	50	100.0	178	3	US-08-828-741B-13

28	50	100.0	178	3	US-09-160-567-13	Sequence 13, Appl
29	50	100.0	178	4	US-09-710-299-13	Sequence 13, Appl
30	50	100.0	178	4	US-09-509-031-13	Sequence 13, Appl
31	50	100.0	197	3	US-08-897-020-7	Sequence 7, Appl
32	50	100.0	197	3	US-09-350-823-7	Sequence 7, Appl
33	50	100.0	198	3	US-08-828-741B-8	Sequence 8, Appl
34	50	100.0	198	3	US-09-160-567-8	Sequence 8, Appl
35	50	100.0	198	4	US-09-710-299-8	Sequence 8, Appl
36	50	100.0	198	4	US-09-509-031-8	Sequence 8, Appl
37	50	100.0	334	4	US-10-017-736C-4	Sequence 6, Appl
38	50	100.0	342	3	US-08-828-741B-6	Sequence 6, Appl
39	50	100.0	342	3	US-09-160-567-6	Sequence 6, Appl
40	50	100.0	342	4	US-09-710-299-6	Sequence 6, Appl
41	50	100.0	342	4	US-09-509-031-6	Sequence 6, Appl
42	50	100.0	386	3	US-08-895-707-7	Sequence 7, Appl
43	50	100.0	409	4	US-10-017-736C-2	Sequence 2, Appl
44	50	100.0	495	3	US-08-828-741B-4	Sequence 4, Appl
45	50	100.0	495	3	US-09-160-567-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-660-626-3
Sequence 3, Application US/08660626
Patent No. 5789655
GENERAL INFORMATION:
APPLICANT: Stanley B. Prusiner
APPLICANT: Glenn C. Telling
APPLICANT: Fred B. Cohen
APPLICANT: Michael R. Scott
TITLE OF INVENTION: TRANSGENIC ANIMALS EXPRESSING
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Ascii
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/660,626
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Valeta Gregg
REGISTRATION NUMBER: 35,127
REFERENCE/DOCKET NUMBER: 07532/003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 322-5070
TELEFAX: (415) 854-0875
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-660-626-3

Query Match 100.0% Score 50; DB 1; Length 9;
Best Local Similarity 87.5%; Pred. No. 4, 1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0;
Gaps 0;
QY 1 WXPQFGG 8
|:|||||
Db 2 WRHPQFG 9

```

RESULT 2
US-08-828-741B-14
; Sequence 14, Application US/08828741B
; Patent No. 6043069
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Sues, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 garden city plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,741B
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Digilio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SAMS UR
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-828-741B-14

Query Match          100.0%; Score 50; DB 3; Length 9;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 WXHPOFGG 8
       |:|||||
       2 WRHPOFGG 9

RESULT 3
US-08-948-097-1
; Sequence 1, Application US/08948097C
; Patent No. 6103493
; GENERAL INFORMATION:
; APPLICANT: Skerra, Arne
; APPLICANT: Voss, Selma
; TITLE OF INVENTION: Streptavidin Mutelins
; FILE REFERENCE: HIBR 1119
; CURRENT APPLICATION NUMBER: US/08/948,097C
; EARLIER FILING DATE: 1997-10-09
; CURRENT APPLICATION NUMBER: DE 196 41 876.3
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 1
; LENGTH: 9
; TYPE: PRT

```

```

; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: BINDING
; OTHER INFORMATION: Binding ligand for streptavidin
US-08-948-097-1

Query Match          100.0%; Score 50; DB 3; Length 9;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 WXHPOFGG 8
       |:|||||
       2 WRHPOFGG 9

RESULT 4
US-09-031-168-3
; Sequence 3, Application US/09031168
; Patent No. 6150583
; GENERAL INFORMATION:
; APPLICANT: Stanley B. Prusiner
; APPLICANT: Glenn C. Telling
; APPLICANT: Fred B. Cohen
; APPLICANT: Michael R. Scott
; TITLE OF INVENTION: TRANSGENIC ANIMALS EXPRESSING
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: California
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Abciti
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,168
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/660,626
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Valeta Gregg
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 07532/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 322-5070
; TELEFAX: (415) 954-0875
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-031-168-3

Query Match          100.0%; Score 50; DB 3; Length 9;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 WXHPOFGG 8
       |:|||||
       2 WRHPOFGG 9

RESULT 5
US-09-160-567-14
; Sequence 14, Application US/09160567

```

Patent No. 6326179
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Sues, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/160,567
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,741
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-160-567-14

Query Match 100.0%; Score 50; DB 3; Length 9;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFGG 8
Db 2 WRHPOFGG 9

RESULT 6
US-09-672-239-3
; Sequence 3, Application US/096722239
; Patent No. 6448377
; GENERAL INFORMATION:
; APPLICANT: Lee, Tae Weon
; APPLICANT: Kobliska, Brian
; TITLE OF INVENTION: MODIFIED G PROTEIN SUBUNITS
; FILE REFERENCE: STAN-204
; CURRENT APPLICATION NUMBER: US/09/672,239
; CURRENT FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-672-239-3

Query Match 100.0%; Score 50; DB 4; Length 9;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFGG 8
Db 2 WRHPOFGG 9

RESULT 7
US-09-710-299-14
; Sequence 14, Application US/09710299
; Patent No. 6521741
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Sues, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/710,299
; FILING DATE: 09-No. 6521741-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,741
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-710-299-14

Query Match 100.0%; Score 50; DB 4; Length 9;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFGG 8
Db 2 WRHPOFGG 9

RESULT 8
US-09-509-031-14
; Sequence 14, Application US/09509031
; Patent No. 6590080
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank

```
/ APPLICANT: Sues, Gabriele M.
/ APPLICANT: Tarlinton, David M.
/ APPLICANT: Treutlein, Herbert R.
/ TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF PRODUCING SAME
/ FILE REFERENCE: 13474
/ CURRENT APPLICATION NUMBER: US/09/509,031
/ NUMBER OF SEQ ID NOS: 16
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 14
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence:Strep-tag
US-09-509-031-14
```

```
Query Match      100.0%; Score 50; DB 4; Length 9;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXHRPOFG 8
       1:|||||
Db      2 WXHRPOFG 9
```

```
RESULT 9
US-09-669-516C-3
/ Sequence 3, Application US/09669516C
/ Patent No. 6602672
/ GENERAL INFORMATION:
/ APPLICANT: Prubiner, Stanley B.
/ APPLICANT: Telling, Glenn C.
/ APPLICANT: Cohen, Fred E.
/ APPLICANT: Sebel, Michael R.
/ TITLE OF INVENTION: RECOMBINANT CONSTRUCT ENCODING EPITOPE
/ FILE REFERENCE: UCAL-045CON
/ CURRENT APPLICATION NUMBER: US/09/669,516C
/ PRIOR FILING DATE: 2000-09-25
/ PRIOR APPLICATION NUMBER: 09/031,168
/ PRIOR FILING DATE: 1998-02-26
/ PRIOR APPLICATION NUMBER: 08/660,626
/ PRIOR FILING DATE: 1996-06-06
/ PRIOR APPLICATION NUMBER: 08/521,992
/ PRIOR FILING DATE: 1995-08-31
/ PRIOR APPLICATION NUMBER: 08/509,261
/ PRIOR FILING DATE: 1995-07-31
/ PRIOR APPLICATION NUMBER: 08/242,188
/ PRIOR FILING DATE: 1994-05-13
/ NUMBER OF SEQ ID NOS: 15
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 3
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Strep peptide
US-09-669-516C-3
```

```
Query Match      100.0%; Score 50; DB 4; Length 9;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXHRPOFG 8
       1:|||||
Db      2 WXHRPOFG 9
```

```
RESULT 10
US-08-294-386C-11
/ Sequence 11, Application US/08294386C
/ Patent No. 5646030
```

```
/ GENERAL INFORMATION:
/ APPLICANT: Ray, Bryan L.
/ APPLICANT: Lin, Edmund C.C.
/ APPLICANT: Crea, Roberto
/ TITLE OF INVENTION: Method of Isolating Mutant Cells
/ NUMBER OF SEQUENCES: 20
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Lappin & Kusmer
/ STREET: 200 State Street
/ CITY: Boston
/ STATE: MA
/ COUNTRY: U.S.A.
/ ZIP: 02109
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/294,386C
/ FILING DATE: August 23, 1994
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Keiner, Ann-Louise
/ REGISTRATION NUMBER: 33,523
/ REFERENCE/DOCKET NUMBER: SYZZ-010CIP
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617/330-1300
/ TELEFAX: 617/330-1311
/ INFORMATION FOR SEQ ID NO: 11:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 10 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ HYPOTHEICAL: NO
US-08-294-386C-11
```

```
Query Match      100.0%; Score 50; DB 1; Length 10;
Best Local Similarity 87.5%; Pred. No. 0.018;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXHRPOFG 8
       1:|||||
Db      3 WXHRPOFG 10
```

```
RESULT 11
US-08-737-316A-5
/ Sequence 5, Application US/08737316A
/ Patent No. 5849576
/ GENERAL INFORMATION:
/ APPLICANT: SKERRA, Arne
/ APPLICANT: WARDENBERG, Christina
/ TITLE OF INVENTION: USE OF THE TETRACYCLINE PROMOTER FOR THE
/ TITLE OF INVENTION: STRINGENTLY REGULATED PRODUCTION OF RECOMBINANT PROTEINS IN PR
/ NUMBER OF SEQUENCES: 8
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Nikaido, Marmelstein, Murray & Oram LLP
/ STREET: 655 Fifteenth Street N.W. Suite 330
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: U.S.A.
/ ZIP: 20005-5701
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: IBM PC compatible
/ SOFTWARE: PC-DOS/MS-DOS
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/737,316A
/ FILING DATE: 12-NOV-1996
```

CLASSIFICATION: 435
PRIOR APPLICATION DATA: PCT/EP95/01862
FILING DATE: 17-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 44 17 598.1
FILING DATE: 19-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: KITEP, Monica C.
REGISTRATION NUMBER: 36,105
REFERENCE/DOCKET NUMBER: P564-6012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-5000
TELEFAX: (202) 638-4810
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-737-316A-5

Query Match 100.0%; Score 50; DB 2; Length 10;
Best Local Similarity 87.5%; Pred. No. 0.018;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPQFG 8
|:|||||
Db 3 WRHPQFG 10

RESULT 12
US-08-897-020-6
Sequence 6, Application US/08897020
Patent No. 6028176
GENERAL INFORMATION:
APPLICANT: Shanafelt, Armen; Greve, Jeffrey; Rocznialk, Steven
TITLE OF INVENTION: High-affinity Interleukin-4 Mutains
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bayer Corporation, Pharmaceutical Division
STREET: 400 Morgan Lane
CITY: West Haven
STATE: CT
COUNTRY: United States of America
ZIP: 06516-4175
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS v. 6.30
SOFTWARE: Word for Windows 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/897,020
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: P-91,242
FILING DATE: 19-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Huw R. Jones
REGISTRATION NUMBER: 33, 916
REFERENCE/DOCKET NUMBER: WH5020
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203) 812-2317
TELEFAX: (203) 812-5492
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 10
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
DESCRIPTION: tag for streptavidin
HYPOTHETICAL: no

ANTI-SENSE: no
US-08-897-020-6

Query Match 100.0%; Score 50; DB 3; Length 10;
Best Local Similarity 87.5%; Pred. No. 0.018;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPQFG 8
|:|||||
Db 3 WRHPQFG 10

RESULT 13
US-08-895-707-11
Sequence 11, Application US/08895707
Patent No. 6077700
GENERAL INFORMATION:
APPLICANT: (Pharmacia & Upjohn, Co.)
APPLICANT: alternatively, for U.S. filing:
APPLICANT: Hollingsworth, Robert A.
APPLICANT: Sharma, Satish K.
APPLICANT: Rank, Kenneth B.
APPLICANT: Evans, David B.
TITLE OF INVENTION: Special Constructs and Complexes of
TITLE OF INVENTION: Cyclin E
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pharmacia & Upjohn Company
STREET: 301 Henrietta Street
CITY: Kalamazoo
STATE: MI
COUNTRY: USA
ZIP: 49001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Rel. #1.0, Ver. #1.25/WordPerfect 5.2+
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/895,707
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Wootton, Thomas A.
REGISTRATION NUMBER: 35,004
REFERENCE/DOCKET NUMBER: 6054
TELECOMMUNICATION INFORMATION:
TELEPHONE: 616-833-7914
TELEFAX: 616-833-8897
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-895-707-11

Query Match 100.0%; Score 50; DB 3; Length 10;
Best Local Similarity 87.5%; Pred. No. 0.018;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPQFG 8
|:|||||
Db 3 WRHPQFG 10

RESULT 14
US-09-350-823-6
Sequence 6, Application US/09350823

```

; Patent No. 6313272
; GENERAL INFORMATION:
; APPLICANT: Stanafelt, Armen; Greve, Jeffrey; Rocznik, Steven
; TITLE OF INVENTION: High-affinity interleukin-4 Muteins
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Bayer Corporation, Pharmaceutical Division
; STREET: 400 Morgan Lane
; CITY: West Haven
; STATE: CT
; COUNTRY: United States of America
; ZIP: 06516-4175
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS v. 6.30
; SOFTWARE: Word for Windows 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/350,823
; FILING DATE: 09-Jul-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/897,020
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Huw R. Jones
; REGISTRATION NUMBER: 33, 916
; REFERENCE/DOCKET NUMBER: WH5020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203) 812-2317
; TELEFAX: (203) 812-5492
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; DESCRIPTION: tag for streptavidin
; HYPOTHEICAL: no
; ANTI-SENSE: no
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-350-823-6
;
Query Match      100.0%; Score 50; DB 3; Length 10;
Best Local Similarity 87.5%; Pred.No. 0.018;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Cy      1 WXHRPQFG 8
        |||||
        3 WXHRPQFG 10

Db
;
RESULT 15
US-09-715-805-13:
; Sequence 13, Application US/09715805
; Patent No. 6716626
; GENERAL INFORMATION:
; APPLICANT: Iton, No. 6716626uyuki
; APPLICANT: Kavanaugh, W. Michael
; TITLE OF INVENTION: HUMAN FGF-21 GENE AND GENE EXPRESSION
; FILE REFERENCE: PP-16758, 001/201130, 408
; CURRENT APPLICATION NUMBER: US/09/715,805
; CURRENT FILING DATE: 2000-11-16
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 10
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Residues which bind to paramagnetic streptavidin
```

```

; OTHER INFORMATION: beads (used for purification).
US-09-715-805-13
Query Match      100.0%; Score 50; DB 4; Length 10;
Best Local Similarity 87.5%; Pred.No. 0.018;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Cy      1 WXHRPQFG 8
        |||||
        3 WXHRPQFG 10

Db
;
Search completed: March 2, 2005, 12:25:32
Job time : 10.1057 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 6.11382 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ2
Perfect score: 48
Sequence: 1 wnhpgek 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	87.5	103	2 E90973	hypothetical prote
2	42	87.5	103	2 A85821	unknown protein en
3	42	87.5	103	2 C85713	unknown protein en
4	42	87.5	103	2 A90904	hypothetical prote
5	39	81.2	485	2 PC4427	Nck, Ash and phosph
6	38	79.2	131	2 S6537	Ig heavy chain V r
7	37	77.1	657	1 XJRTK	transketolase (EC
8	37	77.1	666	2 AF0815	transketolase (EC
9	37	77.1	667	1 A48660	transketolase (EC
10	37	77.1	667	2 G91044	transketolase 2 is
11	37	77.1	667	2 B85889	protein M10D9.2 (i
12	36	75.0	257	2 G86021	hypothetical prote
13	36	75.0	414	2 G84311	conserved hypotet
14	36	75.0	443	2 A82719	endo-1,4-beta-xyla
15	36	75.0	466	2 AE1487	cytolysin vvh pre
16	36	75.0	466	2 AH1126	tolerance to colic
17	36	75.0	471	2 A41478	hydrogenase (EC 1.
18	36	75.0	480	2 F97500	hydrogenase-1 larg
19	36	75.0	597	1 HQECL	hydrogenase-1 larg
20	36	75.0	597	2 AE0721	hydrogenase-1 larg
21	36	75.0	597	2 E85632	hydrogenase-1 larg
22	36	75.0	597	2 A90770	hydrogenase-1 larg
23	36	75.0	705	2 T20278	hypothetical prote
24	36	75.0	922	2 T20277	hypothetical prote
25	36	75.0	1381	2 G83405	hypothetical prote
26	35	72.9	334	2 S35523	DNA-directed DNA p
27	35	72.9	334	2 E90813	DNA polymerase III
28	35	72.9	334	2 A85673	DNA polymerase III
29	35	72.9	363	2 T26842	hypothetical prote

30	35	72.9	499	2 AG0882	probable amino aci
31	35	72.9	670	2 T37483	hypothetical pta-
32	35	72.9	808	2 T14513	hypothetical prote
33	35	72.9	1435	2 D96693	protein Putative A
34	35	72.9	3149	1 Q0BE8	BPL1 protein - hu
35	34	70.8	59	2 A43985	myosin-light-chain
36	34	70.8	141	2 G90762	hypothetical prote
37	34	70.8	141	2 H85625	hypothetical prote
38	34	70.8	141	2 E90913	hypothetical prote
39	34	70.8	207	2 JH0145	hypothetical prote
40	34	70.8	228	2 A31403	membrane protein B
41	34	70.8	272	2 JG4170	trypsin-like prote
42	34	70.8	288	2 T10477	sec3 protein - ye
43	34	70.8	291	2 S61498	chemotactic methyl
44	34	70.8	297	2 T39905	protein transport
45	34	70.8	297	2 A45442	transport versicle

ALIGNMENTS

RESULT 1
E90973
hypothetical protein ECs2757 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #ext_change 09-Jul-2004
C:Accession: E90973
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90973
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <NAV>
A:Cross-references: UNIPROT:O8X4V0; GB:BA000007; PIDD:BA036180.1; PIDD:G13362225; GSPDB:
C:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics: ECs2757
A:Gene: ECs2757

Query Match
Best Local Similarity 62.5%; Pred. No. 1.4;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WNHPEK 8
Db 17 WTHPEYK 24

RESULT 2
A85821
unknown protein encoded within prophage CP-933U [imported] - Escherichia coli (strain O1
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #ext_change 09-Jul-2004
C:Accession: A85821
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
iller, L.; Groth, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potlowski, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85821
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <STO>
A:Cross-references: UNIPROT:O8X4V0; GB:AE005174; NID:G12516136; PIDD:AA057029.1; GSPDB:
C:Experimental source: strain O157:H7, substrain ED1933
C:Genetics:
A:Gene: Z3120

Query Match
Best Local Similarity 87.5%; Score 42; DB 2; Length 103;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPPOFEK 8
|:|:|:|
Db 17 WTHPEYER 24

RESULT 3

C85713
unknown protein encoded within prophage CP-9330 [imported] - Escherichia coli (strain O157:H7)
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C/Accession: C85713
R/Perna, N.T.; Plumett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potomousis, K.; Apodaca, Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: C85713
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-103 <STO>
A/Cross-references: UNIPROT:O8X549; GB:AB005174; NID:q12515046; PIDN:AGS6167.1; GSPDB:G
A/Experimental source: strain O157:H7, substrain EDL933
A/Genetic: Z2097

Query Match 87.5%; Score 42; DB 2; Length 103;
Best Local Similarity 62.5%; Pred. No. 1.4;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPPOFEK 8
|:|:|:|
Db 17 WTHPEYER 24

RESULT 4

A90904
hypothetical protein ECs2201 [imported] - Escherichia coli (strain O157:H7, substrain R1)
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C/Accession: A90904
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gaswara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H. DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene A/Reference number: A99629; MUID:21156231; PMID:1126796
A/Accession: A90904
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-103 <HAY>
A/Cross-references: UNIPROT:O8X549; GB:BA000007; PIDN:BAR35624.1; PID:q13361667; GSPDB:G
A/Experimental source: strain O157:H7, substrain R1MD 0509952
A/Genetic: ECs2201

Query Match 87.5%; Score 42; DB 2; Length 103;
Best Local Similarity 62.5%; Pred. No. 1.4;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPPOFEK 8
|:|:|:|
Db 17 WTHPEYER 24

RESULT 5

PC4427
Nck, Ash and phospholipase C gamma-binding protein 4 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 04-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
C/Accession: PC4427
R/Marucka, K.; Miki, H.; Takahashi, K.; Takenawa, T. Biochem. Biophys. Res. Commun. 239, 488-492, 1997
A/Title: A novel ligand for an SH3 domain of the adaptor protein Nck bears an SH2 domain

A/Reference number: PC4427; MUID:98008866; PMID:9344857
A/Accession: PC4427
A/Status: nucleic acid sequence not shown
A/Molecule type: mRNA
A/Residues: 1-485 <MAT>
A/Cross-references: UNIPROT:O14512; DDBJ:AB005216; NID:q2443366; PIDN:BA22432.1; PID:g-
F/31-43-65-78,180-191,220-230/Region: proline-rich
F/277-414/Domain: SH2 homology <SH2>

Query Match 81.2%; Score 39; DB 2; Length 485;
Best Local Similarity 71.4%; Pred. No. 26;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPPOFEK 7
|:|:|:|
Db 336 WCHPFE 342

RESULT 6

S66537
Ig heavy chain V region - mouse (fragment)
C/Species: Mus musculus (house mouse)
C/Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C/Accession: S66537
R/Tsotis, G.; Haase, W.; Engel, A.; Michel, H. Eur. J. Biochem. 231, 823-830, 1995
A/Title: Isolation and structural characterization of trimeric cyanobacterial photosystem A/Reference number: S66536; MUID:95377318; PMID:7649183
A/Accession: S66537
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-131 <TSI>
A/Cross-references: EMBL:X88902; NID:g995869; PIDN:CAA61364.1; PID:g1103701
F/15-98/Domain: immunoglobulin homology <IMW>

Query Match 79.2%; Score 38; DB 2; Length 131;
Best Local Similarity 83.3%; Pred. No. 9.2;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPPOFEK 6
|:|:|:|
Db 124 WRPPOF 129

RESULT 7

XURTK
transketolase (EC 2.2.1.1) - Rhodobacter sphaeroides
N/Alternate names: glycylaldehydetransferase
C/Species: Rhodobacter sphaeroides
C/Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
C/Accession: B41080
R/Chen, J.H.; Gibson, J.L.; McCue, L.A.; Tabita, F.R. J. Biol. Chem. 266, 20447-20452, 1991
A/Title: Identification, expression, and deduced primary structure of transketolase and A/Reference number: A41080; MUID:92041881; PMID:1939038
A/Accession: B41080
A/Molecule type: DNA
A/Residues: 1-657 <CHR>
A/Cross-references: UNIPROT:P29277; GB:M68914; NID:g151988; PIDN:AAA26155.1; PID:g151980
A/Note: The authors translated the codon GGC for residue 173 as Ala
C/Comment: The active enzyme catalyzes the transfer of a keto group to an aldehyde acceptor.
C/Superfamily: transketolase; thiamin pyrophosphate-binding domain homology
C/Keywords: Calvin cycle, homodimer, magnesium, pentose phosphate pathway; thiamin pyrophosphate-binding domain homology <TPB>
F/147-197/Domain: thiamin pyrophosphate-binding domain homology <TPB>

Query Match 77.1%; Score 37; DB 1; Length 657;
Best Local Similarity 71.4%; Pred. No. 83;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WKHPOPE 7
|:|:|
Db 280 WHHPPE 286

RESULT 8
AF0815
transketolase (EC 2.2.1.1) - Salmonella enterica subsp. enterica serovar Typhi (strain C
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AF0815
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Author: Park, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A>Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; UID:21534947; PMID:11677608
A:Accession: AF0815
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-666 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07704.1; PID:g16503690; GSPDB:GN00176
C:Genetics:
A:Gene: STY271
C:Superfamily: transketolase; thiamin pyrophosphate-binding domain homology
C:Keywords: transketase

Query Match 77.1%; Score 37; DB 2; Length 666;
Best Local Similarity 71.4%; Pred. No. 84;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WKHPOPE 7
|:|:|
Db 278 WHHPPE 284

RESULT 9
A48660
transketolase (EC 2.2.1.1) B - Escherichia coli (strain K-12)
N:Alternate names: glycolaldehydetransferase B; transketolase (EC 2.2.1.1) 2
C:Species: Escherichia coli
C:Date: 16-Feb-1994 #sequence_revision 04-Oct-1996 #text_change 09-Jul-2004
C:Accession: A48660; H55021
R:Rida, A.; Teshiba, S.; Mizubuchi, K.
J. Bacteriol. 175, 5375-5383, 1993
A>Title: Identification and characterization of the tktB gene encoding a second transket
A:Reference number: A48660; UID:93374831; PMID:8396116
A:Accession: A48660
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-667 <IID>
A:Cross-references: UNIPROT:P33570; GB:D12473; NID:g440349; PIDN:BA02039.1; PID:g460975
A:Experimental source: strain K-12
A:Note: sequence extracted from NCBI backbone (NCBIN:137560, NCBI:P:137561)
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A:Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; UID:97426617; PMID:9278503
A:Accession: H65021
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-667 <BLAT>
A:Cross-references: GB:AE000333; GB:U00096; NID:g1788805; PIDN:ACG75518.1; PID:g1788808;
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: tktB
A:Map position: 53 min
C:Complex: homodimer
C:Function:
A:Description: catalyzes the formation of D-ribose 5-phosphate and D-xylulose 5-phosphat

C:Superfamily: transketolase; thiamin pyrophosphate-binding domain homology
C:Keywords: Calvin cycle; homodimer; magnesium; pentose phosphate pathway; thiamin pyrop
F:143-193/Domain: thiamin pyrophosphate-binding domain homology <1PB>

Query Match 77.1%; Score 37; DB 1; Length 667;
Best Local Similarity 71.4%; Pred. No. 84;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WKHPOPE 7
|:|:|
Db 278 WHHPPE 284

RESULT 10
G91044
transketolase 2 isozyme [imported] - Escherichia coli (strain O157:H7, substrain RMD 05
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: G91044
R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; UID:21156231; PMID:11258796
A:Accession: G91044
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-667 <HAY>
A:Cross-references: UNIPROT:O8XB1; GB:BA000007; PIDN:BA036750.1; PID:g13362797; GSPDB:G
A:Experimental source: strain O157:H7, substrain RMD 050952
C:Genetics:
A:Gene: EC03327
C:Superfamily: transketolase; thiamin pyrophosphate-binding domain homology

Query Match 77.1%; Score 37; DB 2; Length 667;
Best Local Similarity 71.4%; Pred. No. 84;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WKHPOPE 7
|:|:|
Db 278 WHHPPE 284

RESULT 11
B85889
transketolase 2 isozyme [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: B85889
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; UID:21074935; PMID:11206551
A:Accession: B85889
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-667 <STO>
A:Cross-references: UNIPROT:O8XB1; GB:AE005174; NID:g12516840; PIDN:ACG57574.1; GSPDB:G
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: tktB
C:Superfamily: transketolase; thiamin pyrophosphate-binding domain homology

Query Match 77.1%; Score 37; DB 2; Length 667;
Best Local Similarity 71.4%; Pred. No. 84;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WKHPOPE 7
|:|:|
Db 278 WHHPPE 284

RESULT 12
G88021
Protein W10D9.2 [imported] - *Caenorhabditis elegans*
C/Species: *Caenorhabditis elegans*
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C/Accession: G88021
R/Anonymous: The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A/Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biology
A/Reference number: A75000; MUID:99069613; PMID:9851916
A/Note: see webstee genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_ele
A/Note: published extra appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and
A/Accession: G88021
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-257 <STO>
A/Cross-references: GB:chr_II; PIDN:AB71056.1; PID:G2429529; GSPDB:GN00020; CESP:W10D9.2
C/Genetics:
A/Gene: W10D9.2
A/Map position: 2

Query Match 75.0%; Score 36; DB 2; Length 257;
Best Local Similarity 62.5%; Pred. No. 44;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 WXHPOFEK 8
DB 157 WGHPLFRK 164

RESULT 13
G84311
Hypothetical protein Vng1585c [imported] - *Halobacterium* sp. NRC-1
C/Species: *Halobacterium* sp. NRC-1
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C/Accession: G84311
R/NG, W.V.; Kennedy, S.P.; Mahatras, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.
Jung, K.H.; Alam, M.; Kellar, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jbdlc
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A/Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li
A/Title: Genome sequence of *Halobacterium* species NRC-1.
A/Reference number: AB4160; MUID:20504483; PMID:11016950
A/Accession: G84311
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-414 <STO>
A/Cross-references: UNIPROT:Q9HPK8; GB:AE004437; NID:G10581067; PIDN:AG19859.1; GSPDB:G
C/Genetics:
A/Gene: VNG1585C

Query Match 75.0%; Score 36; DB 2; Length 414;
Best Local Similarity 62.5%; Pred. No. 75;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 WXHPOFEK 8
DB 141 WMSPEFEK 148

RESULT 14
AB2719
conserved hypothetical protein Atu1155 [imported] - *Agrobacterium tumefaciens* (strain C5
C/Species: *Agrobacterium tumefaciens*
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 12-Jul-2004
C/Accession: AB2719
R/Mood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.B.; Chen, Y.; Woo, L.
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; Mccllell
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: AB2719
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-443 <KUR>
A/Cross-references: UNIPROT:Q8UG85; GB:AE008688; PIDN:AAL42168.1; PID:G17739556; GSPDB:G
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: Atu1155
A/Map position: circular chromosome
C/Superfamily: inner membrane protein Cred

Query Match 75.0%; Score 36; DB 2; Length 443;
Best Local Similarity 71.4%; Pred. No. 81;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 WXHPOFE 7
DB 230 WPHQFE 236

RESULT 15
AE1487
endo-1,4-beta-xylinase homolog lin0436 [imported] - *Listeria innocua* (strain C1p11262)
C/Species: *Listeria innocua*
C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C/Accession: AE1487
R/Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Bagnero, F.; Berche, P.; Bloeker
D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
Science 294, 849-852, 2001
A/Authors: Kreft, U.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maltournam, A.; Ma
ok, C.; Schueller, T.; Simoes, N.; Tlerez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A/Title: Comparative genomics of *Listeria* species.
A/Reference number: AB1077; MUID:21537279; PMID:11679669
A/Accession: AE1487
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-466 <GLA>
A/Cross-references: UNIPROT:Q92EM2; GB:AL592022; PIDN:CAC95669.1; PID:G16412865; GSPDB:G
A/Experimental source: strain C1p11262
C/Genetics:
A/Gene: lin0436

Query Match 75.0%; Score 36; DB 2; Length 466;
Best Local Similarity 62.5%; Pred. No. 85;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 WXHPOFEK 8
DB 326 WDHQFLTK 333
Search completed: March 2, 2005, 12:28:48
Job time: 7.11382 secs

UC Mammaliid; Euleneid; Rodentia; Sciurognathia; Muridae; Murinae; Rattus
OX NCB1_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Nakamura I., Yao Y., Suzuki N.;

RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB097860; BAC44887.1; -
 DR HSP: P30803; 1A2S.
 DR GO: GO:0004383; F:guanylate cyclase activity; IEA.
 DR GO: GO:0016829; F:lyase activity; IEA.
 DR GO: GO:0007242; P:intracellular signaling cascade; IEA.
 DR InterPro: IPR01054; G cyclase.
 DR InterPro: IPR009080; tRNA synthetase.
 DR Pfam: PF00211; Guanylate_cyc_1.
 DR SMART: SM00044; CYC1_1.
 DR PROSITE: PS00452; GUANYLATE_CYCLASES_1; 1.
 DR PROSITE: PS0125; GUANYLATE_CYCLASES_2; 1.
 DR Lyase.
 SQ SEQUENCE 743 AA; 83251 MW; ACP5C5B0982813A CRC64;

Query Match 100.0%; Score 48; DB 2; Length 743;
 Best Local Similarity 87.5%; Pred. No. 5.7;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WXPQPEK 8
 DB 736 WSHQPEK 743

RESULT 3
 MATK_MARSC STANDARD; PRT; 508 AA.
 AC O8WK6; 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 05-JUL-2004 (Rel. 44, Last annotation update)
 DE Maturase K (intron maturase).
 GN Name=matk;
 OS Marathrum schiedeanum.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid1; Malpighiales; Podostemaceae; Marathrum.
 RX NCBI_TaxID=116737;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kita Y., Kato M.;
 RT "Phylogenetic relationships of the aquatic angiosperm family
 Podostemaceae inferred from matk sequence data";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Probably assists in splicing chloroplast group II
 introns (by similarity).
 CC -1- SIMILARITY: Belongs to the intron maturase family 2. Matk
 subfamily.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).

CC -----
 CC EMBL: AB038195; BAB83156.1; -
 CC InterPro: IPR008998; Agglutinin.
 CC InterPro: IPR000442; Intron_maturase2.
 CC InterPro: IPR002866; Matk_N.
 CC Pfam: PF01824; Intron_maturase2; 1.
 CC DR Chloroplast; Matk_N; 1.
 CC DR Chloroplast; mRNA processing.
 SQ SEQUENCE 508 AA; 61342 MW; ECCFSB416B0AB5C3 CRC64;

Query Match 89.6%; Score 43; DB 1; Length 508;
 Best Local Similarity 75.0%; Pred. No. 30;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXPQPEK 8
 DB 736 WSHQPEK 743

DB 188 WSHPSFEK 195

RESULT 4
 Q9B8G0 PRELIMINARY; PRT; 508 AA.
 AC Q9B8G0;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Maturase K.
 GN Name=matk;
 OS Vanroyenella plumosa.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid1; Malpighiales; Podostemaceae; Vanroyenella.
 RX NCBI_TaxID=51609;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kita Y., Kato M.;
 RT "Intrafamilial phylogeny of the Aquatic Angiosperm Podostemaceae
 RL inferred from the Nucleotide Sequences of the matk Gene.";
 RL Plant Biol. 3:156-163(2001).
 DR EMBL: AB048378; BAB33398.1; -
 DR GO: GO:0009507; C:chloroplast; IEA.
 DR GO: GO:0008380; P:RNA splicing; IEA.
 DR InterPro: IPR008998; Agglutinin.
 DR InterPro: IPR000442; Intron_maturase2.
 DR Pfam: PF01824; Intron_maturase2; 1.
 DR Pfam: PF01348; Intron_maturase2; 1.
 DR Chloroplast.
 SQ SEQUENCE 508 AA; 61288 MW; 90229C45138AF4B6 CRC64;

Query Match 89.6%; Score 43; DB 2; Length 508;
 Best Local Similarity 75.0%; Pred. No. 30;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXPQPEK 8
 DB 188 WSHPSFEK 195

RESULT 5
 Q9B8G3 PRELIMINARY; PRT; 508 AA.
 AC Q9B8G3;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
 DE Maturase K.
 GN Name=matk;
 OS Oeerya coulteriana.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid1; Malpighiales; Podostemaceae; Oeerya.
 RX NCBI_TaxID=51602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kita Y., Kato M.;
 RT "Intrafamilial phylogeny of the Aquatic Angiosperm Podostemaceae
 RL inferred from the Nucleotide Sequences of the matk Gene.";
 RL Plant Biol. 3:156-163(2001).
 DR EMBL: AB048375; BAB33395.1; -
 DR GO: GO:0009507; C:chloroplast; IEA.
 DR GO: GO:0008380; P:RNA splicing; IEA.
 DR InterPro: IPR008998; Agglutinin.
 DR InterPro: IPR000442; Intron_maturase2.
 DR InterPro: IPR002866; Matk_N.
 DR Pfam: PF01348; Intron_maturase2; 1.
 DR Pfam: PF01824; Matk_N; 1.

```

KW Chloroplast.
SQ SEQUENCE 508 AA; 61400 MW; 45240C7D31B54CDF CRC64;

Query Match 89.6%; Score 43; DB 2; Length 508;
Best Local Similarity 75.0%; Pred. No. 30;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WKXPOFEK 8
DB 188 WSHPSFEK 195

RESULT 6
08X549 PRELIMINARY; PRT; 103 AA.
ID 08X549 Q7ADV0;
AC 08X549 Q7ADV0;
DT 01-MAR-2002 (TRMBLrel. 20, Created)
DT 01-MAR-2002 (TRMBLrel. 20, Last sequence update)
DE 25-OCT-2004 (TRMBLrel. 28, Last annotation update)
DE Hypothetical protein Z2097 (Hypothetical protein Ec62201).
GN OrderedLocName=Ec62201, Z2097;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxId=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuwara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Reg. 8:11-22(2001).
DR EMBL; AB005346; AAG56167.1; -.
DR EMBL; AP002557; BAB35624.1; -.
DR PIR; A90904; A90904.
DR PIR; C85713; C85713.
DR PIR; A90904; A90904.
DR PIR; C85713; C85713.
GN Hypothetical protein; Complete proteome.
SQ SEQUENCE 103 AA; 11980 MW; 5A42A9F1CF29B6 CRC64;

Query Match 87.5%; Score 42; DB 2; Length 103;
Best Local Similarity 62.5%; Pred. No. 8.1;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKXPOFEK 8
DB 17 WTHPEYEK 24

RESULT 7
08X4V0 PRELIMINARY; PRT; 103 AA.
ID 08X4V0 Q7ACU2;
AC 08X4V0 Q7ACU2;
DT 01-MAR-2002 (TRMBLrel. 20, Created)
DT 01-MAR-2002 (TRMBLrel. 20, Last sequence update)
DE 25-OCT-2004 (TRMBLrel. 28, Last annotation update)
DE Hypothetical protein Z3120 (Hypothetical protein Ec62757).
GN OrderedLocName=Ec62757, Z3120;

```

```

OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxId=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuwara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Reg. 8:11-22(2001).
DR EMBL; AB005422; AAG57029.1; -.
DR EMBL; AP002559; BAB36180.1; -.
DR PIR; A85821; A85821.
DR PIR; E90973; E90973.
GN Hypothetical protein; Complete proteome.
SQ SEQUENCE 103 AA; 12092 MW; D4F3CE5A1089D30 CRC64;

Query Match 87.5%; Score 42; DB 2; Length 103;
Best Local Similarity 62.5%; Pred. No. 8.1;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKXPOFEK 8
DB 17 WTHPEYEK 24

RESULT 8
Q9BBG8 PRELIMINARY; PRT; 508 AA.
ID Q9BBG8;
AC Q9BBG8;
DT 01-JUN-2001 (TRMBLrel. 17, Created)
DT 01-JUN-2001 (TRMBLrel. 17, Last sequence update)
DE 01-MAR-2004 (TRMBLrel. 26, Last annotation update)
DR EMBL; AB004836; BAB33390.1; -.
DR EMBL; AB048367; BAB33390.1; -.
DR GO; GO:0009507; C:chloroplast; IEA.
DR GO; GO:0008380; P:rna splicing; IEA.
DR InterPro; IPR008998; A9glutinin.
DR InterPro; IPR00442; Intron_maturase2.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
GN Chloroplast.

```

SEQ SEQUENCE 508 AA; 61332 MW; 0F1D952D1A104B4 CRC64;
 Query Match 83.3%; Score 40; DB 2; Length 508;
 Best Local Similarity 62.5%; Pred. No. 1e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WKHPPEK 8
 Db 188 WNHPSFOK 195

RESULT 9
 SOC7_HUMAN STANDARD; PRT; 485 AA.
 ID SOC7_HUMAN
 AC 014512;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Suppressor of cytokine signaling 7 (SOCS-6) (Nck, Ash and phospholipase C gamma-binding protein) (Nck-associated protein 4) (NAP-4) (Fragment).
 GN Name=SOCS7; Synonyms=NAP4, SOCS6;
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA MEDLINE=98008866; PubMed=9344857; DOI=10.1006/dbrc.1997.7492;
 RT "A novel ligand for an SH2 domain of the adaptor protein Nck bears an RT domain and nuclear signaling motifs."
 RL Biochem. Biophys. Res. Commun. 239:488-492(1997).
 CC -1- FUNCTION: SOCS family proteins form part of a classical negative feedback system that regulates cytokine signal transduction.
 CC -1- SUBUNIT: Interacts, via the third proline-rich region, with the second SH3 domain of the adapter protein NCK. Also interacts with GRB2 and phospholipase C-gamma.
 CC -1- TISSUE SPECIFICITY: Expressed in brain and leukocytes. Also in fetal lung fibroblasts and fetal brain.
 CC -1- SIMILARITY: Contains 1 SH2 domain.
 CC -1- SIMILARITY: Contains 1 SOCS box domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB005216; BAA22432.1; -;
 DR PIR; PC4427; PC4427.
 DR HSSP; P23727; IOO4.
 DR GeneW; HGNC:29846; SOCS7.
 DR MIM; 608788; -;
 DR GO; GO:0017124; F:SH2-domain binding; NMS.
 DR InterPro; IPR000980; SH2.
 DR InterPro; IPR001496; SOCS_C.
 DR Pfam; PF07525; ClpP; 1.
 DR Pfam; PF00017; SH2; 1.
 DR ProDom; PD000093; SH2; 1.
 DR SMART; SM00252; SH2; 1.
 DR SMART; SM00253; SOCS; 1.
 DR PROSITE; PS50001; SH2; 1.
 DR PROSITE; PS50225; SOCS; 1.
 DR Growth regulation; SH2 domain; Signal transduction inhibitor.
 KW NON TER
 FT 1
 FT DOMAIN 31 81 Pro-rich.
 FT DOMAIN 180 260 Pro-rich.
 FT DOMAIN 277 386 SH2.
 FT DOMAIN 381 431 SOCS box.

FT DOMAIN 18 25 Poly-Gly.
 FT DOMAIN 57 64 Poly-Gln.
 SEQ SEQUENCE 485 AA; 53564 MW; 0CC5EC107174A4F9 CRC64;
 Query Match 81.2%; Score 39; DB 1; Length 485;
 Best Local Similarity 71.4%; Pred. No. 1.5e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WKHPPE 7
 Db 336 WCHPPE 342

RESULT 10
 SOC7_MOUSE STANDARD; PRT; 579 AA.
 ID SOC7_MOUSE
 AC 08VHQ2;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 25-OCT-2004 (Rel. 45, Last sequence update)
 DE Suppressor of cytokine signaling 7.
 GN Name=SOCS7; Synonyms=Csh7, Nap4;
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6;
 RA Hilton D.J., Viney E.M., Alexander W.S., Wilson T.A., Nicola N.A.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBD databases.
 CC -1- FUNCTION: SOCS family proteins form part of a classical negative feedback system that regulates cytokine signal transduction.
 CC -1- SIMILARITY: Contains 1 SH2 domain.
 CC -1- SIMILARITY: Contains 1 SOCS box domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF424814; AAL60516.1; -;
 DR HSSP; P23727; IOO4.
 DR MCD; MG1:265188; SOCS7.
 DR InterPro; IPR000980; SH2.
 DR InterPro; IPR001496; SOCS_C.
 DR Pfam; PF07525; ClpP; 1.
 DR Pfam; PF00017; SH2; 1.
 DR ProDom; PD000093; SH2; 1.
 DR SMART; SM00252; SH2; 1.
 DR SMART; SM00253; SOCS; 1.
 DR PROSITE; PS50001; SH2; 1.
 DR PROSITE; PS50225; SOCS; 1.
 DR Growth regulation; SH2 domain; Signal transduction inhibitor.
 KW NON TER
 FT 82
 FT DOMAIN 301 381 Pro-rich.
 FT DOMAIN 398 507 Pro-rich.
 FT DOMAIN 502 552 SH2.
 FT DOMAIN 502 552 SOCS box.
 FT DOMAIN 84 97 Poly-Pro.
 FT DOMAIN 141 149 Poly-Gly.
 FT DOMAIN 181 185 Poly-Gln.
 FT DOMAIN 186 195 Poly-Pro.
 FT DOMAIN 301 310 Poly-Pro.
 FT DOMAIN 341 348 Poly-Pro.
 SEQ SEQUENCE 579 AA; 62783 MW; FAB66BF2A0B685A CRC64;
 Query Match 81.2%; Score 39; DB 1; Length 579;
 Best Local Similarity 71.4%; Pred. No. 1.8e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPPOE 7
|:|:|:|
Db 457 WCHPKFE 463

RESULT 11

ID Q9FNFS PRELIMINARY; PRT; 1335 AA.

AC Q9FNFS;
DT 01-MAR-2001 (TRMBLrel. 16, Created)
DT 01-MAR-2001 (TRMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TRMBLrel. 22, Last annotation update)
DE Similarity to En/Spm-like transposon protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsia.
OX NCBI_TaxId=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically designed P1 clones."
RL DNA Ref. 4:291-300(1997).
DR EMBL; AB006101; BAB10393.1; -.
DR InterPro; IPR005048; DUF287.
DR Pfam; PF03384; DUF287; 1.
SQ SEQUENCE 1335 AA; 152370 MW; D45FCFF80DD0D5F8 CRC64;

Query Match 1 81.2%; Score 39; DB 2; Length 1335;
Best Local Similarity 71.4%; Pred. No. 4.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPPOE 7
|:|:|:|
Db 209 WXPPOE 215

RESULT 12

ID Q6SOC7 PRELIMINARY; PRT; 127 AA.

AC Q6SOC7;
DT 25-OCT-2004 (TRMBLrel. 28, Created)
DT 25-OCT-2004 (TRMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TRMBLrel. 28, Last annotation update)
DE CrCB protein.
GN Name=CrCB; ORFNames=MS2226;
OS Mannheimia succiniciproducens MBE155E.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Mannheimia.
OX NCBI_TaxId=221988;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MBE155E;
RA Hong S.H., Kim J.S., Lee S.Y., In Y.H., Choi S.S., Rih J.-K.,
RA Kim C.H., Jeong H., Hur C.-G., Kim J.G.;
RT "The genome sequence of the capnophilic rumen bacterium Mannheimia
RT succiniciproducens."
RL Nat. Biotechnol. 0:0-0(2004).
DR EMBL; AB016827; AAU3833.1; -.
SQ SEQUENCE 127 AA; 14089 MW; B844F39F46C38E8C CRC64;

Query Match 1 79.2%; Score 38; DB 2; Length 127;
Best Local Similarity 83.3%; Pred. No. 52;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPPOE 6
|:|:|:|
Db 58 WXPPOE 63

RESULT 13
ID Q7ID58 PRELIMINARY; PRT; 176 AA.

AC Q7ID58;
DT 05-JUL-2004 (TRMBLrel. 27, Created)
DT 05-JUL-2004 (TRMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRMBLrel. 27, Last annotation update)
DE CG13011 (Fragment).
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7245;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22887302; PubMed=14525923; DOI=10.1101/gr.1311003;
RA Domazet-Looso T., Tautz D.;
RT "An evolutionary analysis of orphan genes in Drosophila."
RL Genome Res. 13:2213-2219(2003).
DR EMBL; AF532012; AAQ09911.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR008253; Marvel.
DR Pfam; PF01284; MARVEL; 1.
FT NON_TER 1
SQ SEQUENCE 176 AA; 19490 MW; 33E9FA9F347E899 CRC64;

Query Match 1 79.2%; Score 38; DB 2; Length 176;
Best Local Similarity 83.3%; Pred. No. 74;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPPOE 6
|:|:|:|
Db 130 WXPPOE 135

RESULT 14

ID Q9VXD1 PRELIMINARY; PRT; 176 AA.

AC Q9VXD1;
DT 01-MAY-2000 (TRMBLrel. 13, Created)
DT 01-MAY-2000 (TRMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TRMBLrel. 25, Last annotation update)
DE CG13011-PA
GN ORFNames=CG13011;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celinker S.B., Holt R.A., Evans C.A., Gockyne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Zhang M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borrova D., Botchan M.A., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durkin K.J., Evangelista C.C., Ferraz C., Fertlera S., Fleischmann W.,
RA Foster C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegyan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasco P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,
 RA Palazzolo M., Pitman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Schaefer F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Switek R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Maasatman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodgett, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhou X., Zhou X., Zhu C., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426065; PubMed=12537568;
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoekins R.A., Laverly T., Muzny D.M., Nelson C.R.,
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
 RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
 RT "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
 RT *melanogaster* euchromatic sequence,"
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426070; PubMed=12537573;
 RA Krommiller B., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RT a genomic perspective,"
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22426069; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RT systematic review,"
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RN [5]
 RP SEQUENCE FROM N.A.
 RG Playasee; 1
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RG Playasee;
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB003503; AA048643.1; -;
 DR Playasee; FBgn0030771; CG13011.
 DR GO: GO:0016020; C:membrane; IEA.
 DR InterPro: IPR008253; Marvel.
 DR Pfam: PF01284; MARVEL; 1.
 SQ SEQUENCE 176 AA; 19460 MW; FDCALADBJ316A1D1 CRC64;

Query Match 79.2%; Score 38; DB 2; Length 176;
 Best Local Similarity 83.3%; Pred. No. 74;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WKHPDF 6
 |:|||||

Db 130 WKHPDF 135
 RESULT 15
 ID Q90B81 PRELIMINARY; PRT; 207 AA.
 AC Q90B81;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Nef.
 GN Name=nef;
 OS Human immunodeficiency virus 1.
 OC Viruses; Retrod viruses; Retroviridae; Lentivirus.
 OX NCBI_TaxID=11676;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21637800; PubMed=11779357; DOI=10.1089/088922201753342103;
 RA Washishi T., Loubser S., Hide W., Hunt G., Morris L., Ramjee G.,
 RA Abdool-Karim S., Williamson C., Gray C.M.;
 RT "Conserved domains of subtype C nef from South African HIV type 1-
 RT infected individuals include cytotoxic T lymphocyte epitope-rich
 RT regions,"
 RL AIDS Res. Hum. Retroviruses 17:1681-1687(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Washishi T.N., Loubser S., Hide W., Hunt G., Morris L., Ramjee G.,
 RA Abdool-Karim S., Williamson C., Gray C.,
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF397538; AA09477.1; -;
 DR HSP; P04324; IZEC.
 DR GO: GO:000525; F:GTP binding; IEA.
 DR Pfam: PF00469; F-protein; 1.
 DR ProDom: PD000031; HIV Nef; 1.
 KW AIDS; Lipoprotein; Myristate.
 SQ SEQUENCE 207 AA; 23626 MW; FECT3BEA5D79D78 CRC64;

Query Match 79.2%; Score 38; DB 2; Length 207;
 Best Local Similarity 62.5%; Pred. No. 89;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 WKHPDF 8
 |:|||||

Db 198 WKHPDF 205

Search completed: March 2, 2005, 12:44:10
 Job time: 31.7236 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 36.0325 Seconds
(without alignments) cell updates/sec
85.869 Million

Title: SEQ2
Perfect score: 48
Sequence: 1 wxhpgfek 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1Dec04:*

1: geneeqp1980s:*\n2: geneeqp1990s:*\n3: geneeqp2000s:*\n4: geneeqp2001s:*\n5: geneeqp2002s:*\n6: geneeqp2003as:*\n7: geneeqp2003bs:*\n8: geneeqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	8	2	AAW59212 Streptavi
2	48	100.0	8	4	AAE35433 Epitope p
3	48	100.0	8	4	AAE35442 Nascent p
4	48	100.0	8	4	AAE68616 Strept-Tag
5	48	100.0	8	5	AAO19065 Mutation
6	48	100.0	8	5	ABBS7464 STRP tag
7	48	100.0	8	5	ABBS7486 AAV-helpe
8	48	100.0	8	6	ABG73584 Streptagi
9	48	100.0	8	6	AAE37230 Strept-tag
10	48	100.0	8	6	ABP60361 Streptavi
11	48	100.0	8	6	ADA09808 Streptavi
12	48	100.0	8	7	ADA09808 Streptavi
13	48	100.0	8	7	ADA84588 Streptavi
14	48	100.0	8	7	ADB85500 Streptavi
15	48	100.0	8	7	ADD29930 Antibody-
16	48	100.0	8	8	ADDO6951 Streptag
17	48	100.0	8	8	ADP90832 Protein/p
18	48	100.0	8	8	ADP90832 Strept tag
19	48	100.0	8	8	ADR72509 dehydroge
20	48	100.0	9	5	ABG31054 Peptide p
21	48	100.0	10	4	AAI97662 Influenza
22	48	100.0	10	4	AAE99027 Streptavi
23	48	100.0	10	5	AAU80475 Peptide S
24	48	100.0	10	6	ABP56623 C-termina
25	48	100.0	10	8	ADFI1064 Strept-tag

26	48	100.0	10	8	ADN16967 Human res
27	48	100.0	10	8	ADO26489 Strept-tag
28	48	100.0	11	6	AAE38373 Epitope t
29	48	100.0	11	8	ADS20228 Strept tag
30	48	100.0	19	6	ABG74881 Bacteriop
31	48	100.0	19	6	ABG74882 Bacteriop
32	48	100.0	19	6	ABG74880 Bacteriop
33	48	100.0	19	8	AD139157 Construct
34	48	100.0	19	8	AD139155 Construct
35	48	100.0	19	8	AD139156 Construct
36	48	100.0	24	6	ABP60363 Streptavi
37	48	100.0	24	6	ABP60362 Streptavi
38	48	100.0	35	6	ABP60369 Streptavi
39	48	100.0	36	6	ABP60370 Streptavi
40	48	100.0	117	5	AAU97558 Synthetic
41	48	100.0	117	5	AAU97553 Synthetic
42	48	100.0	117	5	AAU97557 Synthetic
43	48	100.0	117	5	AAU97559 Synthetic
44	48	100.0	117	5	AAU97555 Synthetic
45	48	100.0	117	5	AAU97556 Synthetic

ALIGNMENTS

RESULT 1
AAW59212 standard; peptide; 8 AA.
ID AAW59212;

27-AUG-1998 (first entry)

Streptavidin tagged peptide ligand #2.

Streptavidin; ligand; binding affinity; mutant; isolation; purification; recover; immobilise.

Synthetic.

EP835934-A2.

15-APR-1998.

09-OCT-1997; 97BP-00117504.

10-OCT-1996; 96DE-01041876.

(BIOA-) INST BIOANALYTIK GMBH.

Skerza A, Voss S;

WPI; 1998-218868/20.

Streptavidin mutants with higher binding affinity for peptide ligands - have mutation in amino acid region 44-53, used to isolate, purify or determine fusion proteins including these ligands.

Claim 10; Page 11; 21BP; German.

AAW59211 and AAW59212 are ligands used in a method to assay binding affinity of streptavidin mutants. These mutants have a mutation within the amino acid (aa) region 44-53 of the wild-type protein show a higher binding affinity than the wild-type for peptide ligands that include the sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin mutants can be used to isolate, purify and determine proteins or to determine/recover substances that contain streptavidin-binding groups. Such compounds may also be used to immobilise fusions on microtitre plates, microbeads or sensor chips.

Sequence 8 AA;

SO

Query Match 100.0%; Score 48; DB 2; Length 8;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXHPOFEK 8
 Db 1 WSHPOFEK 8

RESULT 2

AAB35433
 ID AAB35433 standard; peptide; 8 AA.

AC AAB35433;

DT 23-MAY-2001 (first entry)

DE Eptlope peptide #3.

KW Naescent protein detection; protein analysis; aminoacylated tRNA;
 BODIPY marker; disease diagnosis.

OS Unidentified.

PN WO200114578-A1.

PD 01-MAR-2001.

PF 23-AUG-2000; 2000WO-US023233.

PR 25-AUG-1999; 99US-00382736.

PR 25-AUG-1999; 99US-00382950.

PA (AMBE-) AMBERGEN INC.

PI Rothschild KJ, Gite S, Olejnik J;

DR WPI; 2001-168972/17.

PT Method for detecting nascent proteins by fluorescence comprises
 misaminoacylating a tRNA molecule with a marker compound, useful for
 detecting mutations in proteins, e.g. cancer.

PS Disclosure; Page 47; 204pp; English.

CC The present invention describes a method of detecting nascent proteins
 involving aminoacylating a tRNA molecule with a 4',4-difluoro-4-bora-3A,4A
 -diaz-a-indacene (BODIPY) marker leading to the production of a
 CC misaminoacylated tRNA. This enables the detection, isolation and analysis
 CC of nascent proteins using UV without the usual accompanying radioactivity
 CC problems. It may be used to detect mutations, for example in cancer,
 CC Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer
 CC
 SQ Sequence 8 AA;

Query Match

Best Local Similarity 100.0%; Score 48; DB 4; Length 8;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXHPOFEK 8
 Db 1 WSHPOFEK 8

RESULT 3

AAB35442
 ID AAB35442 standard; peptide; 8 AA.

AC AAB35442;

DT 23-MAY-2001 (first entry)

DB Naescent protein detection method related peptide #4.

XX
 KW Naescent protein detection; protein analysis; aminoacylated tRNA;
 KW BODIPY marker; disease diagnosis.

OS Unidentified.

PN WO200114578-A1.

PD 01-MAR-2001.

PF 23-AUG-2000; 2000WO-US023233.

PR 25-AUG-1999; 99US-00382736.

PR 25-AUG-1999; 99US-00382950.

PA (AMBE-) AMBERGEN INC.

PI Rothschild KJ, Gite S, Olejnik J;

DR WPI; 2001-168972/17.

PT Method for detecting nascent proteins by fluorescence comprises
 misaminoacylating a tRNA molecule with a marker compound, useful for
 detecting mutations in proteins, e.g. cancer.

PS Example 22; Page 153; 204pp; English.

CC The present invention describes a method of detecting nascent proteins
 involving aminoacylating a tRNA molecule with a 4',4-difluoro-4-bora-3A,4A
 -diaz-a-indacene (BODIPY) marker leading to the production of a
 CC misaminoacylated tRNA. This enables the detection, isolation and analysis
 CC of nascent proteins using UV without the usual accompanying radioactivity
 CC problems. It may be used to detect mutations, for example in cancer,
 CC Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer
 CC
 SQ Sequence 8 AA;

Query Match

Best Local Similarity 100.0%; Score 48; DB 4; Length 8;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXHPOFEK 8
 Db 1 WSHPOFEK 8

RESULT 4

AAB68616
 ID AAB68616 standard; peptide; 8 AA.

AC AAB68616;

DT 27-APR-2001 (first entry)

DE Strep-Tag II sequence.

KW Protein-RNA fusion; Strep-Tag II.

OS Unidentified.

PN WO200107657-A1.

PD 01-FEB-2001.

PF 19-JUL-2000; 2000WO-US019653.

PR 27-JUL-1999; 99US-0145834P.

PA (PHYL-) PHYLOS INC.

PI Kurz M, Lohse P, Wagner R;

DR WPI; 2001-182803/18.

```

XX Affixing a peptide acceptor to an RNA molecule useful for producing
PT fusion proteins for isolating proteins or nucleic acids with desired
PT properties through attachment of a peptide acceptor to the 3' end of an
PT RNA molecule.
XX
XX Example 5; Page 22; 56pp; English.
XX
CC The present invention relates to a method for affixing a peptide acceptor
CC to an RNA molecule through the formation of a covalent bond, noncovalent
CC bond, or by chemical ligation. The method is useful for producing RNA-
CC protein fusions which can be used for the isolation of proteins or
CC nucleic acids with desired properties from large pools of partially or
CC completely random amino acid or nucleic acid sequences. The present
CC sequence is a Strip-Tag II used in the present invention
XX
SQ Sequence 8 AA;

Query Match 100.0%; Score 48; DB 4; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.8e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXPQPEK 8
   |||||
Db 1 WSHQPEK 8

RESULT 5
AA019065
ID AA019065 standard; peptide; 8 AA.
XX
AC AA019065;
XX
DT 14-NOV-2002 (first entry)
XX
DE Mutation detection method tag peptide SEQ ID NO: 34.
XX
KM Mutation detection; primer; mutant; tag; tumour suppressor gene;
KM protein production; cancer.
XX
OS Synthetic.
XX
PN WO20026675-A2.
XX
PD 29-AUG-2002.
XX
PF 15-FEB-2002; 2002WO-EP001651.
XX
PR 16-FEB-2001; 2001DE-01007317.
XX
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Kahmann S, Mueller O;
XX
PI MPI: 2002-674859/72.
XX
DR N-PSDB; AAL49464.
XX
PT Detecting mutations in nucleic acid, useful for diagnosis and
PT characterization of tumors, by amplification, in vitro transcription and
PT translation, then protein detection.
XX
PS Disclosure; Fig 5; 62pp; German.
XX
XX The present invention relates to a method of detecting mutations in a
XX nucleic acid by amplifying the nucleic acid to produce a double-stranded
XX amplicon, in vitro transcription and translation of this amplicon, and
XX detection of the translated protein. The primers used for amplification
XX are designed to produce an amplicon that is translatable and allows
XX differentiation between translation products of wild-type and mutated
XX nucleic acids. The method is used to detect mutations in tumour
XX suppressor genes, for (early) diagnosis, monitoring and characterisation
XX of tumours (especially of bladder and intestines) and in the germ line
XX (using nucleic acids from embryos or blood cells). A new multi-tag vector

```

```

CC is used to detect or verify the reading frame of a nucleic acid cloned in
CC it, and to determine the suitability of detectable peptides for analysis
CC and/or purification of a recombinant protein, expressed from a sequence
CC cloned in the vector. The present sequence is a tag peptide which was
CC used in the invention
XX
SQ Sequence 8 AA;

Query Match 100.0%; Score 48; DB 5; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.8e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 WXPQPEK 8
   |||||
Db 1 WSHQPEK 8

RESULT 6
ABB57464
ID ABB57464 standard; peptide; 8 AA.
XX
AC ABB57464;
XX
DT 18-MAR-2002 (first entry)
XX
DE STRIP tag II peptide.
XX
KM Immunomodulatory human MHC class II antigen-binding protein; HLA;
KM human leukocyte antigen; immune system; immunosuppression; antibody;
KM major histocompatibility complex; antirheumatic; antiarthritic;
KM neuroprotective; antiinflammatory; antidiabetic; antipsoriatic;
KM immunosuppressive; dermatological; antithyroid; nephrotropic; psoriasis;
KM chymotryptic; hepatotropic; immune response suppressor; narcolepsy;
KM rheumatoid arthritis; juvenile arthritis; multiple sclerosis; insulin;
KM Grave's disease; insulin-dependent diabetes; Hashimoto's disease;
KM systemic lupus erythematosus; ankylosing spondylitis; myasthenia gravis;
KM transplant rejection; graft versus host disease; pemphigus vulgaris;
KM glomerulonephritis; thyroiditis; pancreatitis; primary biliary cirrhosis;
KM irritable bowel disease; Sjogren's syndrome.
XX
OS Synthetic.
XX
PN WO200187338-A1.
XX
PD 22-NOV-2001.
XX
PF 14-MAY-2001; 2001WO-US015626.
XX
PR 12-MAY-2000; 2000EP-00110063.
XX
PR 06-OCT-2000; 2000US-0238762P.
XX
PA (GPCB-) GPC BIOTECH AG.
XX
PA (MORP-) MORPHOSYS AG.
XX
PI Nagy Z, Tesar M, Thomassen-Wolf B;
XX
PI MPI: 2002-075289/10.
XX
DR Composition for suppressing immune response, treating diseases of immune
XX system, has polypeptide comprising antibody-based antigen-binding domain
XX of human composition, which binds antigen expressed on a cell surface.
XX
PS Disclosure; Page 31; 139pp; English.
XX
XX The present invention describes a composition (I), comprising a
XX polypeptide comprising an antibody-based antigen-binding domain of human
XX composition with binding specificity for an antigen expressed on the
XX surface of a cell, where treating cells expressing the antigen with the
XX polypeptides leads to suppression of an immune response, and the IC50 for
XX the suppression of immune response is 1 microm or less. (I) has
XX antirheumatic, antiarthritic, neuroprotective, antiinflammatory,
XX antidiabetic, antipsoriatic, immunosuppressive, dermatological,
XX antithyroid, nephrotropic, chymotryptic and hepatotropic activities, and

```

CC can be used as a suppressor of immune response. (1) is useful for
 CC suppressing activation or proliferation of a cell of the immune system,
 CC suppressing IL-2 secretion by a cell, the interaction of a cell of the
 CC immune system with another cell, immunosuppressing a patient and for
 CC killing a cell expressing an antigen, human leukocyte antigen (HLA)-DR on
 CC the surface of the cell, where neither cytotoxic antigen nor
 CC immunological mechanisms are needed to cause or lead to the killing. (1)
 CC optionally linked to cytotoxic or immunogenic agent) is useful for
 CC preparing a pharmaceutical preparation for the treatment of rheumatoid
 CC arthritis, juvenile arthritis, multiple sclerosis, Grave's disease,
 CC insulin-dependent diabetes, narcolepsy, psoriasis, systemic lupus,
 CC erythematous, ankylosing spondylitis, transplant rejection, graft versus
 CC host disease, Hashimoto's disease, myasthenia gravis, pemphigus vulgaris,
 CC glomerulonephritis, thyroiditis, pancreatitis, insulin, primary biliary
 CC cirrhosis, irritable bowel disease and Sjogren's syndrome in humans.
 CC ABA92469 to ABA92474 and ABB57457 to ABB57590 represent sequence used in
 CC the exemplification of the present invention
 CC
 SQ Sequence 8 AA;

Query Match 100.0%; Score 48; DB 5; Length 8;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFEK 8
 DB 1 WSHPOFEK 8

RESULT 7
 ABB77486
 ID ABB77486 standard; peptide; 8 AA.
 AC ABB77486;
 XX
 DT 22-JUL-2002. (first entry)

XX AAV-helper plasmid related Strep-peptide tag SEQ ID NO 17.
 DE
 KM Adeno-associated virus; AAV; glycoprotein; ITR; cap; rep; vector.

OS Synthetic.
 PN WO200238782-A2.
 XX
 PD 16-MAY-2002.

PF 13-NOV-2001; 2001WO-EP013125.
 XX
 PR 13-NOV-2000; 2000DE-01056210.

PA (ARIM-) ARIMEDES BIOTECHNOLOGY GMBH.
 PI Orberger G, Hellmuth K, Wagnen C;
 XX
 DR WPI; 2002-435853/46.

PT Vector system for preparing recombinant adeno-associated viral particles,
 PT used for high-level expression of heterologous therapeutic proteins in
 PT eukaryotic cells.
 XX
 PS Disclosure; Page 17; 59pp; German.

CC The invention relates to a viral vector system for preparing recombinant
 CC adeno-associated virus (AAV) particles comprising; at least two plasmid
 CC vectors (ABLS8983 and ABLS8984) that include the two inverted terminal
 CC repeats (ITR) of AAV and additional sequences and plasmid vectors without
 CC ITRs but containing the rep and cap genes of AAV required for replication
 CC and packaging. The system is useful for producing recombinant AAV for
 CC production of a wide range of therapeutic glycoproteins in eukaryotic
 CC cells. The system provides efficient, large scale production of
 CC heterologous proteins in mammalian cells, without requiring an adenovirus
 CC helper. It is not toxic to host cells and does not cause lysis, so

CC produced proteins are highly pure. The present sequence is that of a
 CC peptide tag, useful to the invention
 CC
 SQ Sequence 8 AA;

Query Match 100.0%; Score 48; DB 5; Length 8;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFEK 8
 DB 1 WSHPOFEK 8

RESULT 8
 ABB73584
 ID ABB73584 standard; protein; 8 AA.
 AC ABB73584;
 XX
 DT 03-MAR-2003 (first entry)

XX StreptagII affinity peptide.
 DE
 XX Non-protein L-amino acid; O-acetyl-L-serine; agrochemical production;
 KM nucleophile; O-acetyl-L-serine sulphydrylase; pharmaceutical production.
 KW
 XX Synthetic.
 OS
 XX EPI247869-A1.
 PN
 PD 09-OCT-2002.

PE 28-MAR-2002; 2002EP-00007262.
 XX
 PR 04-APR-2001; 2001DE-01016881.

PR 03-MAY-2001; 2001DE-01021515.
 XX
 PA (CONE) CONSORTIUM ELEKTROCHEM IND GMBH.

XX Maier T, Gaebert C;
 PI
 DR WPI; 2003-077522/08.
 XX

PT Production of non-protein L-amino acids useful for the manufacture of
 PT pharmaceuticals and agrochemicals, comprises an enzyme catalyzed reaction
 PT of O-acetyl-L-serine with a nucleophile.
 XX
 PS Example 1; Page 8; 20pp; German.

CC This invention describes a novel method for the production of non-protein
 CC L-amino acids by the reaction of O-acetyl-L-serine with a nucleophile in
 CC the presence of a catalyst comprising O-acetyl-L-serine sulphydrylase at
 CC pH 5-7.4. The method of the invention is useful for the manufacture of
 CC pharmaceuticals and agrochemicals. In contrast to the process described
 CC in DE 10046934, a high nucleophile concentration can be used which
 CC includes toxic compounds. This sequence represents an affinity peptide
 CC containing a StreptagII motif which is used in the construction of fusion
 CC genes containing Escherichia coli cyxK and cyxM fragments
 CC
 SQ Sequence 8 AA;

Query Match 100.0%; Score 48; DB 6; Length 8;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFEK 8
 DB 1 WSHPOFEK 8

RESULT 9
 AAE37230

ID AAE37230 standard; peptide; 8 AA.
 XX
 AC AAE37230;
 XX
 DT 07-AUG-2003 (first entry)
 XX
 DE Strep-tag II epitope peptide.
 XX
 KW Gene expression; therapy; isolation; epitope.
 XX
 OS Synthetic.
 XX
 PN WO2003038049-A2.
 XX
 PD 08-MAY-2003.
 XX
 PF 29-OCT-2002; 2002WO-US034645.
 XX
 PR 29-OCT-2001; 2001US-0340689P.
 XX
 PA (RENO-) RENOVIS INC.
 XX
 PI Heintz N, Serafini TA, Shyjan AW;
 XX
 DR WPI; 2003-430512/40.
 XX
 PT Isolating cell-type specific mRNAs, useful in gene expression analysis or
 PT quantification in a specific cell in a heterogeneous cell mixture, by
 PT isolating mRNA bound to ribosomes or proteins that bind mRNA in a cell
 PT type specific manner.
 XX
 PS Example 2; page 121; 136pp; English.
 XX
 CC The invention relates to a method for isolating mRNA from a population of
 CC cells. The method involves selectively isolating ribosomes or proteins
 CC that bind mRNA in a cell type specific manner and then isolating the mRNA
 CC bound to the ribosomes or proteins that bind mRNA. The method is useful
 CC for facilitating the analysis and quantification of gene expression in a
 CC selected cell type present within a heterogeneous cell mixture. The
 CC method may also be used in diagnostics or therapies for human diseases.
 CC The present sequence is Strep-tag II epitope peptide. This sequence is
 CC used to illustrate the method of the invention
 CC
 SQ Sequence 8 AA;
 XX
 Query Match 100.0%; Score 48; DB 6; Length 8;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WXHPQFEK 8
 DB 1 WSHPOFEK 8

XX
 PR 21-MAR-2001; 2001DE-01013776.
 XX
 PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
 XX
 PI Schmidt T;
 XX
 DR WPI; 2003-031166/03.
 XX
 PT New isolated peptide, useful as affinity purification tag for recombinant
 PT protein, comprises at least two high-affinity streptavidin-binding
 PT modules.
 XX
 PS Example 1; Page 8; 18pp; German.
 XX
 CC The invention relates to an isolated peptide (I) comprising at least two
 CC individual modules separated by 0-50 amino acids, with each containing at
 CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
 CC streptavidin binding modules, are useful as affinity handles for
 CC purification of recombinant fusion proteins (FP), also for detecting FP,
 CC e.g. on protein chips or microtitre plates. The modules in (I) bind
 CC strongly to streptavidin, with a co-operative effect that provides
 CC stronger binding than a single tag, but are displaced by a competitor.
 CC (I) does not interfere with the function of attached proteins (II) (so it
 CC may not be essential to remove it); facilitates detection and has easily
 CC controllable binding properties. (I) is particularly used for purifying
 CC FP from dilute solution in batch formats (which use simpler apparatus
 CC than column methods and result in lower loss of FP). The present sequence
 CC is that of a streptavidin tag peptide disclosed with the invention
 CC
 SQ Sequence 8 AA;
 XX
 Query Match 100.0%; Score 48; DB 6; Length 8;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WXHPQFEK 8
 DB 1 WSHPOFEK 8

RESULT 11
 ABP60368
 ID ABP60368 standard; peptide; 8 AA.
 XX
 AC ABP60368;
 XX
 DT 28-MAR-2003 (first entry)
 XX
 DE Streptavidin binding peptide SEQ ID NO 9.
 XX
 KW Streptavidin; protein chip; microtitre plate; detection.
 XX
 OS Synthetic.
 XX
 PN DE10113776-A1.
 XX
 PD 02-OCT-2002.
 XX
 PF 21-MAR-2001; 2001DE-01013776.
 XX
 PR 21-MAR-2001; 2001DE-01013776.
 XX
 PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
 XX
 PI Schmidt T;
 XX
 DR WPI; 2003-031166/03.
 XX
 PT New isolated peptide, useful as affinity purification tag for recombinant
 PT protein, comprises at least two high-affinity streptavidin-binding
 PT modules.
 XX

PS Claim 5; Page 16; 18pp; German.
 CC The invention relates to an isolated peptide (I) comprising at least two
 CC individual modules separated by 0-50 amino acids, with each containing at
 CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
 CC streptavidin binding modules, are useful as affinity handles for
 CC purification of recombinant fusion proteins (FP), also for detecting FP,
 CC e.g. on protein chips or microtitre plates. The modules in (I) bind
 CC strongly to streptavidin, with a co-operative effect that provides
 CC stronger binding than a single tag, but are displaced by a competitor.
 CC (I) does not interfere with the function of attached proteins (II) (so it
 CC may not be essential to remove it); facilitates detection and has easily
 CC controllable binding properties. (I) is particularly used for purifying
 CC FP from dilute solution in batch formats (which use simpler apparatus
 CC than column methods and result in lower loss of FP). The present sequence
 CC is that of a streptavidin binding peptide disclosed with the invention
 SQ Sequence 8 AA;
 Query Match 100.0%; Score 48; DB 6; Length 8;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WSHPOPEK 8
 DB 1 WSHPOPEK 8
 RESULT 12
 ID ADB09808 standard; peptide; 8 AA.
 AC ADB09808;
 DT 06-NOV-2003 (first entry)
 DE Streptag epitope useful as an affinity marker.
 XX
 DE Streptag epitope useful as an affinity marker.
 KM Non-radioactive marker; nascent protein detection method;
 KM cellular translation system; cell-free translation system;
 KM dipyrrometheneboron difluoride dye;
 KM 4,4-difluoro-4-bora-3a,4a-diaza-8-indacene dye; vaccine; drug;
 KM human disease screening; human disorder; protein separation;
 KM affinity marker; Streptag epitope.
 OS Synthetic.
 XX
 PN US2003092031-A1.
 XX
 PD 15-MAY-2003.
 XX
 PF 18-JUN-2002; 2002US-00174368.
 XX
 PR 25-AUG-1999; 99US-00382736.
 PR 23-AUG-2000; 2000WO-US023233.
 PR 21-JUN-2002; 2002US-00049332.
 XX
 PA (AMBE-) AMBERGEN INC.
 XX
 PI Rochschild KJ, Gite S, Olejnik J;
 PT WPI; 2003-576764/54.
 DR
 XX
 XX
 PT Detecting, analyzing or isolating nascent proteins comprises introducing
 PT a modified nucleic acid template into a cellular or cell-free translation
 PT system to generate a nascent protein having at least an N-terminal
 PT marker.
 XX
 PS Disclosure; Page 15; 76pp; English.
 XX The present invention relates to non-radioactive markers used in the
 CC detection and analysis of nascent proteins translated in cellular or cell
 CC -free translation systems. The preferred non-radioactive markers are

CC dipyrrometheneboron difluoride (4,4-difluoro-4-bora-3a,4a-diaza-8-
 CC indacene) dyes. The detection method of the invention is a gel-free
 CC method that comprises introducing a modified nucleic acid template into a
 CC translation system under conditions such that a nascent protein is
 CC generated, the protein comprising at least an N-terminal marker. The
 CC method is useful in detecting, analyzing and isolating nascent proteins
 CC produced in a cell-free or cellular translation system without the use of
 CC radioactive amino acids or other radioactive labels. Compositions
 CC comprising nascent proteins translated in the presence of markers may be
 CC used as vaccines or as drugs for humans and other animals. The method and
 CC used as a rapid means for the detection of nascent proteins may be
 CC used as a rapid means to screen humans or other animals for the presence
 CC of certain diseases or disorders. The present sequence represents an
 CC epitope that can be used as an affinity marker for protein separation.
 SQ Sequence 8 AA;
 Query Match 100.0%; Score 48; DB 7; Length 8;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WSHPOPEK 8
 DB 1 WSHPOPEK 8
 RESULT 13
 ID ADB84588 standard; peptide; 8 AA.
 AC ADB84588;
 DT 04-DEC-2003 (first entry)
 DE Streptavidin conserved peptide #2.
 XX
 DE Streptavidin conserved peptide #2.
 KM cell-free transcription system; cell-free translation system;
 KM protein synthesis; matrix; streptavidin.
 OS Escherichia coli.
 XX
 PN DE10137792-A1.
 XX
 PD 27-FEB-2003.
 XX
 PF 06-AUG-2001; 2001DE-01037792.
 XX
 PR 06-AUG-2001; 2001DE-01037792.
 XX
 PA (ERDM/) ERDMANN V.
 XX
 PI Erdmann VA, Lamja T, Stiege W;
 PT WPI; 2003-343999/33.
 DR
 XX
 XX
 PT Expressing genes in cell-free system, useful for preparation of proteins,
 PT comprises that the protein formed is removed from solution by binding to
 PT a matrix.
 XX
 PS Claim 13; Col 8; 8pp; German.
 CC
 CC This invention describes a novel method of expressing genes in a cell-
 CC free transcription and translation system which comprises using a
 CC reaction solution containing all necessary components of the
 CC transcription/translation system, amino acids, nucleotides and
 CC metabolites that supply energy and that are needed for synthesis. The
 CC proteins formed are immobilised on a matrix. The method allows simple
 CC recovery of proteins without a separate isolation step and the amount of
 CC proteins produced can be determined before a reaction is complete.
 CC Continuous removal of proteins prevents it interfering with the
 CC expression process, making possible synthesis of proteins that interact
 CC adversely with the process, so normally produced only in very low yields.
 CC The use of a matrix concentrates the proteins produced and is applicable

CC to proteins of any size. This sequence represents a highly conserved
 CC peptide from E. coli streptavidin which is used to illustrate the method
 CC of the invention.

XX Sequence 8 AA;

Query Match 100.0%; Score 48; DB 7; Length 8;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXHPOFEK 8
 ||:|||||
 Db 1 WSHPOFEK 8

RESULT 14

ID ADB85500 standard; peptide; 8 AA.

AC ADB85500;

DT 04-DEC-2003; (first entry)

DE Streptavidin tag peptide related to human aggrecanase.

XX aggrecanase; aggrecan; articular cartilage; proteoglycan;
 KW cartilage compressibility; cartilage elasticity; arthritic disease;
 KW osteoarthritis; cartilage degradation; inflammatory joint disease;
 KW aggrecanase-1; ADAMTS-4; aggrecanase-2; ADAMTS-5; thrombospondin domain;
 KW TSP domain; osteopathic; antiarthritic; cytosolic; antiinflammatory;
 KW antirheumatic; ophthalmological; thrombolytic; vasodilator; antimicrobial;
 KW respiratory-gen; nocrotropic; neuroprotective; antiparkinsonian;
 KW immunosuppressive; aggrecanase inhibition; cancer; rheumatoid arthritis;
 KW septic arthritis; corneal ulceration; coronary thrombosis;
 KW Croun's disease; emphysema; Alzheimer's disease; Parkinson's disease;
 KW multiple sclerosis; aortic aneurysm; streptavidin tag.

OS Synthetic.

XX WO2003066823-A2.

PD 14-AUG-2003.

XX 05-FEB-2003; 2003WO-US003554.

XX 05-FEB-2002; 2002US-0354592P.

XX (AMHP) WYETH.

XX Georgiadis K, Crawford TK, Tomkinson KM, Morris EA, Racie L;

DR WPI; 2003-73195/69.
 DR N-PSDB; ADB85498.

PT New biologically-active aggrecanase protein having a deletion of all, or
 PT a portion of a TSP domain, useful for treating osteoarthritis, cancer,
 PT Parkinson's disease, coronary thrombosis, Alzheimer's disease and
 PT multiple sclerosis.

XX Example 1; Fig 18; 11pp; English.

XX This invention relates to novel truncated human aggrecanase proteins and
 CC nucleotide sequences. Aggrecan is a major extracellular component of
 CC articular cartilage. It is a proteoglycan responsible for providing
 CC cartilage with its mechanical properties of compressibility and
 CC elasticity. The loss of aggrecan has been implicated in the degradation
 CC of articular cartilage in arthritic diseases such as osteoarthritis.
 CC Aggrecanase is responsible for the cleavage of aggrecan, thereby having a
 CC role in cartilage degradation associated with osteoarthritis and
 CC inflammatory joint disease. The proteins of the current invention are
 CC truncated aggrecanase-1 (ADAMTS-4) and aggrecanase-2 (ADAMTS-5) enzymes
 CC which have at least one thrombospondin (TSP) domain deleted. These are
 CC biologically active and have greater stability and higher expression than

CC their full-length counterparts. The proteins of the invention may be of
 CC use in the development of compounds with osteopathic, antiarthritic,
 CC cytostatic, antiinflammatory, antirheumatic, ophthalmological,
 CC thrombolytic, vasotropic, antimicrobial, respiratory-gen, nocrotropic,
 CC neuroprotective, antiparkinsonian or immunosuppressive activities through
 CC aggrecanase inhibition. The proteins of the invention may therefore be
 CC useful for the manufacture of compositions for the treatment of
 CC aggrecanase-associated conditions, such as osteoarthritis, cancer,
 CC inflammatory joint disease, rheumatoid arthritis, septic arthritis,
 CC corneal ulceration, coronary thrombosis, Croun's disease, emphysema,
 CC Alzheimer's disease, Parkinson's disease, multiple sclerosis and aortic
 CC aneurysm. The present sequence is that of a streptavidin tag peptide
 CC linker which was used during the exemplification of the invention.

XX Sequence 8 AA;

Query Match 100.0%; Score 48; DB 7; Length 8;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXHPOFEK 8
 ||:|||||
 Db 1 WSHPOFEK 8

RESULT 15

ID ADD29930 standard; peptide; 8 AA.

AC ADD29930;

DT 15-JAN-2004 (first entry)

DE Antibody-presenting hollow protein nanoparticle-related peptide #2.

XX drug composition; hollow protein nanoparticle; cell specific antibody;
 KW cancer; virus infection; cytostatic; virucide.

XX Unidentified.

XX WO2003082330-A1.

PD 09-OCT-2003.

XX 26-MAR-2003; 2003WO-JP003694.

XX 29-MAR-2002; 2002JP-00097424.

XX 21-FEB-2003; 2003JP-00045088.

XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX Kuroda S, Tanizawa K, Kondo A, Ueda M, Seno M, Okajima T;

DR WPI; 2003-853719/79.

PT Hollow protein nanoparticles enclosing a drug substance and presenting a
 PT tissue or cell specific antibody at their surface for site-specific
 PT therapy of cancer and other diseases.

XX Example C; SEQ ID NO 28; 96pp; Japanese.

XX The invention comprises a drug composition consisting of hollow protein
 CC nanoparticles which enclose a drug substance and which present a cell
 CC specific antibody on their surface. The drug composition is useful for
 CC cell or tissue specific treatment of cancer and virus infections. The
 CC present amino acid sequence was used in the exemplification of the
 CC invention.

XX Sequence 8 AA;

Query Match 100.0%; Score 48; DB 7; Length 8;
 Best Local Similarity 87.5%; Pred. No. 1.8e+06;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	WXHPQFEK	8
		:	
Db	1	WSHPQFEK	8

Search completed: March 2, 2005, 13:02:44
Job time : 38.0325 secs

GenCore version: 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 24.878 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ2

Perfect score: 48
Sequence: 1 wxhpgfek 8

Scoring table: BLOSUM62DX
Gapop 10.0, Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	8	9	US-09-809-517A-9
2	48	100.0	8	9	US-09-973-145-7
3	48	100.0	8	12	US-09-813-197-8
4	48	100.0	8	13	US-10-104-218-5
5	48	100.0	8	13	US-10-208-357-9
6	48	100.0	8	14	US-10-001-934-8
7	48	100.0	8	14	US-10-026-578B-2
8	48	100.0	8	14	US-10-026-578B-9
9	48	100.0	8	14	US-10-174-368A-7
10	48	100.0	8	14	US-10-264-127-8
11	48	100.0	8	15	US-10-339-712-8
12	48	100.0	8	15	US-10-339-712-67
13	48	100.0	8	15	US-10-275-046-4

14	48	100.0	8	15	US-10-425-000-76	Sequence 76, Appl
15	48	100.0	8	15	US-10-424-999-25	Sequence 25, Appl
16	48	100.0	8	15	US-10-358-283-23	Sequence 23, Appl
17	48	100.0	8	16	US-10-628-432-41	Sequence 41, Appl
18	48	100.0	8	17	US-10-494-248-17	Sequence 17, Appl
19	48	100.0	8	17	US-10-634-645-11	Sequence 11, Appl
20	48	100.0	8	17	US-10-719-523-8	Sequence 8, Appl
21	48	100.0	9	9	US-09-983-067-3	Sequence 3, Appl
22	48	100.0	10	9	US-09-809-517A-6	Sequence 6, Appl
23	48	100.0	10	15	US-10-147-211A-20	Sequence 20, Appl
24	48	100.0	11	15	US-10-354-983-29	Sequence 29, Appl
25	48	100.0	11	16	US-10-628-432-25	Sequence 25, Appl
26	48	100.0	21	9	US-09-809-517A-30	Sequence 30, Appl
27	48	100.0	22	9	US-09-809-517A-33	Sequence 33, Appl
28	48	100.0	24	9	US-09-809-517A-31	Sequence 31, Appl
29	48	100.0	24	14	US-10-026-578B-3	Sequence 3, Appl
30	48	100.0	25	14	US-10-026-578B-4	Sequence 4, Appl
31	48	100.0	25	9	US-09-809-517A-34	Sequence 34, Appl
32	48	100.0	36	14	US-10-026-578B-10	Sequence 10, Appl
33	48	100.0	36	14	US-10-026-578B-11	Sequence 11, Appl
34	48	100.0	117	10	US-09-977-137A-4	Sequence 4, Appl
35	48	100.0	117	10	US-09-977-137A-5	Sequence 5, Appl
36	48	100.0	117	10	US-09-977-137A-7	Sequence 7, Appl
37	48	100.0	117	10	US-09-977-137A-8	Sequence 8, Appl
38	48	100.0	117	10	US-09-977-137A-9	Sequence 9, Appl
39	48	100.0	117	10	US-09-977-137A-10	Sequence 10, Appl
40	48	100.0	117	10	US-09-977-137A-11	Sequence 11, Appl
41	48	100.0	117	10	US-09-977-137A-12	Sequence 12, Appl
42	48	100.0	118	10	US-09-977-137A-6	Sequence 6, Appl
43	48	100.0	245	17	US-10-887-228A-1	Sequence 1, Appl
44	48	100.0	246	17	US-10-887-228A-9	Sequence 9, Appl
45	48	100.0	252	17	US-10-887-228A-5	Sequence 5, Appl

ALIGNMENTS

```
RESULT 1
US-09-809-517A-9
; Sequence 9, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-9

Query Match      100.0%; Score 48; DB 9; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 WXHPOFEK 8
|:|||||
Db 1 WSHPOFEK 8

RESULT 2
US-09-973-145-7

```
; Sequence 7, Application US/09973145
; Patent No. US20020132248A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
; FILE REFERENCE: AMBER-06819
; CURRENT APPLICATION NUMBER: US/09/973,145
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 09/382,950
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; NAME/KEY: misc feature
; OTHER INFORMATION: Synthetic
US-09-973-145-7
```

```
Query Match      100.0%; Score 48; DB 9; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXHPOFEK 8
        |:|||||
Db      1 WSHPOFEK 8
```

```
RESULT 3
US-09-813-197-8
; Sequence 8, Application US/09813197
; Publication No. US2005009013A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Proteins
; FILE REFERENCE: AMBER-03951
; CURRENT APPLICATION NUMBER: US/09/813,197
; CURRENT FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-813-197-8
```

```
Query Match      100.0%; Score 48; DB 12; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXHPOFEK 8
        |:|||||
Db      1 WSHPOFEK 8
```

```
RESULT 4
US-10-104-218-5
; Sequence 5, Application US/10104218
; Publication No. US20020177196A1
; GENERAL INFORMATION:
; APPLICANT: Maier, Thomas
; APPLICANT: Gabbert, Carsten
; TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
```

```
; FILE REFERENCE: MAIER, T. ET AL.-2
; CURRENT APPLICATION NUMBER: US/10/104,218
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 16 881.0
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 21 515.0
; PRIOR FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(8)
; OTHER INFORMATION: StreptagII affinity peptide for protein purification
US-10-104-218-5
```

```
Query Match      100.0%; Score 48; DB 13; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXHPOFEK 8
        |:|||||
Db      1 WSHPOFEK 8
```

```
RESULT 5
US-10-208-357-9
; Sequence 9, Application US/10208357
; Publication No. US20020182687A1
; GENERAL INFORMATION:
; APPLICANT: Kurz, Markus
; APPLICANT: Lohse, Peter
; APPLICANT: Magnet, Richard
; TITLE OF INVENTION: Peptide Acceptor Ligation Methods
; FILE REFERENCE: 50036/031002
; CURRENT APPLICATION NUMBER: US/10/208,357
; CURRENT FILING DATE: 2002-07-30
; PRIOR APPLICATION NUMBER: US/09/619,103
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/145,834
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: designed sequence to act as an identifying tag
US-10-208-357-9
```

```
Query Match      100.0%; Score 48; DB 13; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXHPOFEK 8
        |:|||||
Db      1 WSHPOFEK 8
```

```
RESULT 6
US-10-001-934-8
; Sequence 8, Application US/10001934
; Publication No. US20030032782A1
; GENERAL INFORMATION:
; APPLICANT: Nagy, Zoltan
; TITLE OF INVENTION: HUMAN PEPTIDES/PROTEINS CAUSING OR LEADING TO THE
; KILLING OF INVENTION: KILLING OF CELLS INCLUDING LYMPHOID TUMOR CELLS
; FILE REFERENCE: GRCG-P01-003
; CURRENT APPLICATION NUMBER: US/10/001,934
```

;; CURRENT FILING DATE: 2001-11-15
;; NUMBER OF SEQ ID NOS: 63
;; SOFTWARE: PatentIn Ver. 2.1
;; SEQ ID NO 8
;; LENGTH: 8
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-001-934-8

Query Match 100.0%; Score 48; DB 14; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPQFEK 8
|:|||||
Db 1 WSHPOFEK 8

RESULT 7
US-10-026-578B-2
;; Sequence 2, Application US/10026578B
;; Publication No. US20030083474A1
;; GENERAL INFORMATION:
;; APPLICANT: IBA (GmbH)
;; APPLICANT: Schmidt, Thomas
;; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
;; FILE REFERENCE: 100810.01US1
;; CURRENT APPLICATION NUMBER: US/10/026,578B
;; CURRENT FILING DATE: 2002-11-11
;; PRIOR APPLICATION NUMBER: DE 101 13 776.1
;; PRIOR FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: PCT/EP01/11846
;; PRIOR FILING DATE: 2001-10-12
;; NUMBER OF SEQ ID NOS: 14
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 2
;; LENGTH: 8
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Synthetic Peptide
;; NAME/KEY: misc.feature
;; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-2

Query Match 100.0%; Score 48; DB 14; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPQFEK 8
|:|||||
Db 1 WSHPOFEK 8

RESULT 8
US-10-026-578B-9
;; Sequence 9, Application US/10026578B
;; Publication No. US20030083474A1
;; GENERAL INFORMATION:
;; APPLICANT: IBA (GmbH)
;; APPLICANT: Schmidt, Thomas
;; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
;; FILE REFERENCE: 100810.01US1
;; CURRENT APPLICATION NUMBER: US/10/026,578B
;; CURRENT FILING DATE: 2002-11-11
;; PRIOR APPLICATION NUMBER: DE 101 13 776.1
;; PRIOR FILING DATE: 2001-03-21
;; PRIOR APPLICATION NUMBER: PCT/EP01/11846
;; PRIOR FILING DATE: 2001-10-12

;; NUMBER OF SEQ ID NOS: 14
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 9
;; LENGTH: 8
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Synthetic Peptide
;; NAME/KEY: misc.feature
;; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-9

Query Match 100.0%; Score 48; DB 14; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPQFEK 8
|:|||||
Db 1 WSHPOFEK 8

RESULT 9
US-10-174-368A-7
;; Sequence 7, Application US/10174368A
;; Publication No. US20030092031A1
;; GENERAL INFORMATION:
;; APPLICANT: Rothschild, Kenneth J.
;; APPLICANT: Gite, Sadanand
;; APPLICANT: Olejnik, Jerzy
;; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prc
;; FILE REFERENCE: AMBR-07145
;; CURRENT APPLICATION NUMBER: US/10/174,368A
;; CURRENT FILING DATE: 2002-06-18
;; PRIOR APPLICATION NUMBER: 10/049,332
;; PRIOR FILING DATE: 2002-06-21
;; PRIOR APPLICATION NUMBER: PCT/US00/23233
;; PRIOR FILING DATE: 2000-08-23
;; PRIOR APPLICATION NUMBER: 09/382,736
;; PRIOR FILING DATE: 1999-08-25
;; NUMBER OF SEQ ID NOS: 18
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 7
;; LENGTH: 8
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Synthetic
;; NAME/KEY: misc.feature
;; OTHER INFORMATION: Synthetic
US-10-174-368A-7

Query Match 100.0%; Score 48; DB 14; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPQFEK 8
|:|||||
Db 1 WSHPOFEK 8

RESULT 10
US-10-264-127-8
;; Sequence 8, Application US/10264127
;; Publication No. US20030190643A1
;; GENERAL INFORMATION:
;; APPLICANT: Rothschild, Kenneth
;; APPLICANT: Gite, Sadanand
;; APPLICANT: Olejnik, Jerzy
;; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot
;; FILE REFERENCE: AMBR-03951
;; CURRENT APPLICATION NUMBER: US/10/264,127

```
/ CURRENT FILING DATE: 2002-10-03
/ PRIOR APPLICATION NUMBER: US/09/382,736B
/ PRIOR FILING DATE: 1999-08-25
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 8
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Synthetic
US-10-264-127-8
```

```
Query Match      100.0%; Score 48; DB 14; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      1 WXHPOFEK 8
Db      1 WSHPOFEK 8
```

```
RESULT 11
US-10-339-712-8
/ Sequence 8, Application US/10339712
/ Publication No. US20040014071A1
/ GENERAL INFORMATION:
/ APPLICANT: Rothschild, Kenneth J.
/ APPLICANT: Gite, Sadanand
/ APPLICANT: Olejnik, Jerzy
/ APPLICANT: Lim, Mark
/ TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent
/ FILE OF INVENTION: Proteins
/ FILE REFERENCE: AMBER-07199
/ CURRENT APPLICATION NUMBER: US/10/339,712
/ PRIOR FILING DATE: 2003-01-10
/ PRIOR APPLICATION NUMBER: 10/049,322
/ PRIOR FILING DATE: 2002-06-21
/ PRIOR APPLICATION NUMBER: PCT/US00/23233
/ PRIOR FILING DATE: 2000-08-23
/ PRIOR APPLICATION NUMBER: 09/382,736
/ PRIOR FILING DATE: 1999-08-25
/ NUMBER OF SEQ ID NOS: 84
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 8
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Synthetic
US-10-339-712-8
```

```
Query Match      100.0%; Score 48; DB 15; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      1 WXHPOFEK 8
Db      1 WSHPOFEK 8
```

```
RESULT 12
US-10-339-712-67
/ Sequence 67, Application US/10339712
/ Publication No. US20040014071A1
/ GENERAL INFORMATION:
/ APPLICANT: Rothschild, Kenneth J.
/ APPLICANT: Gite, Sadanand
/ APPLICANT: Olejnik, Jerzy
/ APPLICANT: Lim, Mark
/ TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent
/ FILE OF INVENTION: Proteins
/ FILE REFERENCE: AMBER-07199
```

```
/ CURRENT APPLICATION NUMBER: US/10/339,712
/ CURRENT FILING DATE: 2003-01-10
/ PRIOR APPLICATION NUMBER: 10/049,322
/ PRIOR FILING DATE: 2002-06-21
/ PRIOR APPLICATION NUMBER: PCT/US00/23233
/ PRIOR FILING DATE: 2000-08-23
/ PRIOR APPLICATION NUMBER: 09/382,736
/ PRIOR FILING DATE: 1999-08-25
/ NUMBER OF SEQ ID NOS: 84
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 67
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Synthetic
US-10-339-712-67
```

```
Query Match      100.0%; Score 48; DB 15; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      1 WXHPOFEK 8
Db      1 WSHPOFEK 8
```

```
RESULT 13
US-10-275-046-4
/ Sequence 4, Application US/10275046
/ Publication No. US20040019187A1
/ GENERAL INFORMATION:
/ APPLICANT: Nagy et al.
/ TITLE OF INVENTION: IMMUNOMODULATORY HUMAN MHC CLASS II ANTIGEN-BINDING POLYPEPTIDES
/ FILE REFERENCE: GPCG-P01-260
/ CURRENT APPLICATION NUMBER: US/10/275,046
/ CURRENT FILING DATE: 2002-10-31
/ PRIOR APPLICATION NUMBER: 00110063.5
/ PRIOR FILING DATE: 2000-05-12
/ NUMBER OF SEQ ID NOS: 97
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 4
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: STREP tag II
US-10-275-046-4
```

```
Query Match      100.0%; Score 48; DB 15; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      1 WXHPOFEK 8
Db      1 WSHPOFEK 8
```

```
RESULT 14
US-10-425-000-76
/ Sequence 76, Application US/10425000
/ Publication No. US20040052777A1
/ GENERAL INFORMATION:
/ APPLICANT: Nesbitt, Mark
/ APPLICANT: Cameron, Beatrice
/ APPLICANT: Blanche, Francis
/ TITLE OF INVENTION: Kinkade Polypeptides and Methods for Using Them to Inhibit
/ FILE OF INVENTION: Angiogenesis
/ FILE REFERENCE: ST01027-B
/ CURRENT APPLICATION NUMBER: US/10/425,000
/ CURRENT FILING DATE: 2003-04-29
/ PRIOR APPLICATION NUMBER: 10/233,675
/ PRIOR FILING DATE: 2002-09-04
```

```
/ NUMBER OF SEQ ID NOS: 105
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO: 76
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Purification tag
US-10-425-000-76
```

```
Query Match          100.0%; Score 48; DB 15; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXHPOFEK 8
        |||||
Db       1 WSHPOFEK 8
```

RESULT 15

```
US-10-424-999-25
/ Sequence 25, Application US/10424999
/ Publication No. US20040052810A1
/ GENERAL INFORMATION:
/ APPLICANT: Nesbit, Mark
/ APPLICANT: Cameron, Beatrice
/ APPLICANT: Blanche, Francis
/ TITLE OF INVENTION: Abrogen Polypeptides, Nucleic Acids Encoding Them and Methods for
/ FILE REFERENCE: ST01027-A
/ CURRENT APPLICATION NUMBER: US/10/424,999
/ PRIOR APPLICATION NUMBER: 10/233,675
/ PRIOR FILING DATE: 2002-09-04
/ NUMBER OF SEQ ID NOS: 70
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO: 25
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Purification tag
US-10-424-999-25
```

```
Query Match          100.0%; Score 48; DB 15; Length 8;
Best Local Similarity 87.5%; Pred. No. 1.2e+06;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXHPOFEK 8
        |||||
Db       1 WSHPOFEK 8
```

```
Search completed: March 2, 2005, 14:16:49
Job time : 24.878 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 9.10569 Seconds
(without alignments)
65.585 Million cell updates/sec

Title: SEQ2
Perfect score: 48
Sequence: 1 wkhpfek 8

Scoring table: BLOSUM62DX
gapop 10.0, Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents AA:
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6CTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	8	3	US-08-948-097-2
2	48	100.0	8	3	US-09-382-950-7
3	48	100.0	8	3	US-09-382-736B-8
4	48	100.0	8	4	US-09-619-103-9
5	48	100.0	8	4	US-10-104-218-5
6	48	100.0	8	4	US-09-809-517A-9
7	48	100.0	10	4	US-09-809-517A-30
8	48	100.0	21	4	US-09-809-517A-6
9	48	100.0	22	4	US-09-809-517A-33
10	48	100.0	24	4	US-09-809-517A-31
11	48	100.0	25	4	US-09-809-517A-34
12	48	100.0	117	4	US-09-977-137A-5
13	48	100.0	117	4	US-09-977-137A-4
14	48	100.0	117	4	US-09-977-137A-7
15	48	100.0	117	4	US-09-977-137A-8
16	48	100.0	117	4	US-09-977-137A-9
17	48	100.0	117	4	US-09-977-137A-10
18	48	100.0	117	4	US-09-977-137A-11
19	48	100.0	117	4	US-09-977-137A-12
20	48	100.0	118	4	US-09-977-137A-6
21	40	83.3	951	4	US-09-252-991A-26766
22	39	81.2	207	3	US-09-302-766-36
23	38	79.2	8	3	US-08-948-097-16
24	38	79.2	9	1	US-08-660-626-3
25	38	79.2	9	3	US-08-828-741B-14
26	38	79.2	9	3	US-08-948-097-1
27	38	79.2	9	3	US-09-031-168-3

28	38	79.2	9	3	US-09-160-567-14	Sequence 14, Appl
29	38	79.2	9	4	US-09-672-239-3	Sequence 3, Appl
30	38	79.2	9	4	US-09-710-299-14	Sequence 14, Appl
31	38	79.2	9	4	US-09-509-031-14	Sequence 14, Appl
32	38	79.2	9	4	US-09-669-516C-3	Sequence 3, Appl
33	38	79.2	10	1	US-08-294-386C-11	Sequence 11, Appl
34	38	79.2	10	2	US-08-737-316A-5	Sequence 5, Appl
35	38	79.2	10	3	US-08-897-020-6	Sequence 6, Appl
36	38	79.2	10	3	US-08-895-707-11	Sequence 11, Appl
37	38	79.2	10	3	US-09-350-823-6	Sequence 6, Appl
38	38	79.2	10	4	US-09-715-805-13	Sequence 13, Appl
39	38	79.2	10	4	US-09-692-945-9	Sequence 9, Appl
40	38	79.2	10	4	US-09-640-041-7	Sequence 7, Appl
41	38	79.2	10	5	PCT-US95-10224-11	Sequence 11, Appl
42	38	79.2	12	1	US-08-737-757-17	Sequence 17, Appl
43	38	79.2	12	3	US-08-218-369-8	Sequence 8, Appl
44	38	79.2	12	4	US-09-904-599A-8	Sequence 8, Appl
45	38	79.2	12	5	PCT-US95-03742-8	Sequence 8, Appl

ALIGNMENTS

```

RESULT 1
US-08-948-097-2
Sequence 2, Application US/08948097C
Patent No. 6103493
GENERAL INFORMATION:
APPLICANT: Skeira, Arne
APPLICANT: Voss, Selma
TITLE OF INVENTION: Streptavidin Mureins
FILE REFERENCE: HUBB 1119
CURRENT APPLICATION NUMBER: US/08/948,097C
EARLIER FILING DATE: 1997-10-09
EARLIER APPLICATION NUMBER: DE 196 41 876.3
NUMBER OF SEQ ID NOS: 17
SEQ ID NO 2
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial sequence
FEATURE:
NAME/KEY: BINDING
OTHER INFORMATION: Binding ligand for streptavidin
US-08-948-097-2

Query Match      100.0%; Score 48; DB 3; Length 8;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      1 WKHPOFEK 8
      1:|||||
      2:|||||
      3:|||||
      4:|||||
      5:|||||
      6:|||||
      7:|||||
      8:|||||
      9:|||||
      10:|||||
      11:|||||
      12:|||||
      13:|||||
      14:|||||
      15:|||||
      16:|||||
      17:|||||
      18:|||||
      19:|||||
      20:|||||
      21:|||||
      22:|||||
      23:|||||
      24:|||||
      25:|||||
      26:|||||
      27:|||||
      28:|||||
      29:|||||
      30:|||||
      31:|||||
      32:|||||
      33:|||||
      34:|||||
      35:|||||
      36:|||||
      37:|||||
      38:|||||
      39:|||||
      40:|||||
      41:|||||
      42:|||||
      43:|||||
      44:|||||
      45:|||||

RESULT 2
US-09-382-950-7
Sequence 7, Application US/09382950
Patent No. 6303337
GENERAL INFORMATION:
APPLICANT: Rothschild, Kenneth
APPLICANT: Gite, Sedana
APPLICANT: Olejnik, Jerzy
TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
FILE REFERENCE: AMBER-03879
CURRENT APPLICATION NUMBER: US/09/382,950
CURRENT FILING DATE: 1999-08-25
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.0
SEQ ID NO 7
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial/Unknown

```

```
FEATURE:
NAME/KEY: misc feature
LOCATION: (1..7)
OTHER INFORMATION: Synthetic
US-09-382-950-7
```

Query Match 100.0%; Score 48; DB 3; Length 8;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXPQPEK 8
|:|||||
Db 1 WSHQPEK 8

```
RESULT 3
US-09-382-736B-8
Sequence 8, Application US/09382736B
Patent No. 6306628
GENERAL INFORMATION:
APPLICANT: Rothschild, Kenneth
APPLICANT: Gite, Sadanand
APPLICANT: Olejnik, Jerzy
TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Pro
FILE REFERENCE: AMBER-03951
CURRENT APPLICATION NUMBER: US/09/382,736B
CURRENT FILING DATE: 1999-08-25
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin version 3.0
SEQ ID NO 8
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-09-382-736B-8
```

Query Match 100.0%; Score 48; DB 3; Length 8;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXPQPEK 8
|:|||||
Db 1 WSHQPEK 8

```
RESULT 4
US-09-619-103-9
Sequence 9, Application US/09619103
Patent No. 6429300
GENERAL INFORMATION:
APPLICANT: Kurtz, Markus
APPLICANT: Lohse, Peter
APPLICANT: Wagner, Richard
TITLE OF INVENTION: Peptide Acceptor Ligation Methods
FILE REFERENCE: 50036/031002
CURRENT APPLICATION NUMBER: US/09/619,103
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 60/145,834
PRIOR FILING DATE: 1999-07-27
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: designed sequence to act as an identifying tag
US-09-619-103-9
```

Query Match 100.0%; Score 48; DB 4; Length 8;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXPQPEK 8
|:|||||
Db 1 WSHQPEK 8

```
RESULT 5
US-10-104-218-5
Sequence 5, Application US/10104218
Patent No. 6579705
GENERAL INFORMATION:
APPLICANT: MAIER, Thomas
APPLICANT: GAEBERT, Carsten
TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
FILE REFERENCE: MAIER, T. ET AL.-2
CURRENT APPLICATION NUMBER: US/10/104,218
CURRENT FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: German No. 6579705 101 16 881.0
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: German No. 6579705 101 21 515.0
PRIOR FILING DATE: 2001-05-03
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patentin version 3.0
SEQ ID NO 5
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)..(8)
OTHER INFORMATION: StreptagII affinity peptide for protein purification
US-10-104-218-5
```

Query Match 100.0%; Score 48; DB 4; Length 8;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXPQPEK 8
|:|||||
Db 1 WSHQPEK 8

```
RESULT 6
US-09-809-517A-9
Sequence 9, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: Patentin version 3.0
SEQ ID NO 9
LENGTH: 8
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-9
```

Query Match 100.0%; Score 48; DB 4; Length 8;
Best Local Similarity 87.5%; Pred. No. 4.1e+05;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXPQPEK 8
|:|||||

Db 1 WSHPOFEK 8

RESULT 7

US-09-809-517A-6
; Sequence 6, Application US/09809517A
; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

; FILE REFERENCE: MORPHO/11

; CURRENT APPLICATION NUMBER: US/09/809,517A

; PRIOR FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: EP 99114072.4

; PRIOR FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: EP 00103551.8

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 6

; LENGTH: 10

; TYPE: PRT

; ORGANISM: artificial sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-6

Query Match

Best Local Similarity 100.0%; Score 48; DB 4; Length 10;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFEK 8

Db 3 WSHPOFEK 10

RESULT 8

US-09-809-517A-30
; Sequence 30, Application US/09809517A
; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

; FILE REFERENCE: MORPHO/11

; CURRENT APPLICATION NUMBER: US/09/809,517A

; PRIOR FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: EP 99114072.4

; PRIOR FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: EP 00103551.8

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 30

; LENGTH: 21

; TYPE: PRT

; ORGANISM: artificial sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-30

Query Match

Best Local Similarity 100.0%; Score 48; DB 4; Length 21;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFEK 8

Db 14 WSHPOFEK 21

RESULT 9

US-09-809-517A-33

; Sequence 33, Application US/09809517A

; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

; FILE REFERENCE: MORPHO/11

; CURRENT APPLICATION NUMBER: US/09/809,517A

; PRIOR FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: EP 99114072.4

; PRIOR FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: EP 00103551.8

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 33

; LENGTH: 22

; TYPE: PRT

; ORGANISM: artificial sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-33

Query Match

Best Local Similarity 100.0%; Score 48; DB 4; Length 22;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFEK 8

Db 15 WSHPOFEK 22

RESULT 10

US-09-809-517A-31
; Sequence 31, Application US/09809517A
; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

; FILE REFERENCE: MORPHO/11

; CURRENT APPLICATION NUMBER: US/09/809,517A

; PRIOR FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: EP 99114072.4

; PRIOR FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: EP 00103551.8

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 31

; LENGTH: 24

; TYPE: PRT

; ORGANISM: artificial sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-31

Query Match

Best Local Similarity 100.0%; Score 48; DB 4; Length 24;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFEK 8

Db 17 WSHPOFEK 24

RESULT 11

US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

```

; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34
```

```
Query Match          100.0%; Score 48; DB 4; Length 25;
Best Local Similarity 87.5%; Pred. No. 0.04;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXHPOFEK 8
        |:|||||
Db       18 WSHPOFEK 25
```

```

RESULT 12
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4
```

```
Query Match          100.0%; Score 48; DB 4; Length 117;
Best Local Similarity 87.5%; Pred. No. 0.21;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXHPOFEK 8
        |:|||||
Db       110 WSHPOFEK 117
```

```

RESULT 13
US-09-977-137A-5
; Sequence 5, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
```

```

; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5
```

```
Query Match          100.0%; Score 48; DB 4; Length 117;
Best Local Similarity 87.5%; Pred. No. 0.21;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXHPOFEK 8
        |:|||||
Db       110 WSHPOFEK 117
```

```

RESULT 14
US-09-977-137A-7
; Sequence 7, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-7
```

```
Query Match          100.0%; Score 48; DB 4; Length 117;
Best Local Similarity 87.5%; Pred. No. 0.21;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXHPOFEK 8
        |:|||||
Db       110 WSHPOFEK 117
```

```

RESULT 15
US-09-977-137A-8
; Sequence 8, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8

Query Match 100.0%; Score 48; DB 4; Length 117;
Best Local Similarity 87.5%; Pred. No. 0.21;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 wxhPOPEK 8
||:|||||
DB 110 WSHPOPEK 117

Search completed: March 2, 2005, 12:25:32
Job time : 9.10569 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 6.11382 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ3

Perfect score: 48

Sequence: 1 wxhpgfer 8

Scoring table:

BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 segs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR79:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	81.2	103	2 E90973	hypothetical prote
2	39	81.2	103	2 A85821	unknown protein en
3	39	81.2	103	2 C85713	unknown protein en
4	39	81.2	103	2 A90904	hypothetical prote
5	39	81.2	485	2 PCA427	Nck, Ash and phosp
6	39	81.2	705	2 T20278	hypothetical prote
7	39	81.2	922	2 T20277	hypothetical prote
8	38	79.2	131	2 S65337	Ig heavy chain V r
9	37	77.1	657	1 XJRTFK	transketolase (EC
10	37	77.1	666	2 AF0815	transketolase (EC
11	37	77.1	667	1 A48660	transketolase (EC
12	37	77.1	667	2 G91044	transketolase 2 is
13	37	77.1	667	2 B85889	transketolase 2 is
14	36	75.0	313	2 H69297	conserved cytochrom
15	36	75.0	433	2 H83444	probable cytochrom
16	36	75.0	443	2 AB2719	conserved hypothet
17	36	75.0	471	2 A41478	cytolysin vrbn pre
18	36	75.0	480	2 P97500	tolerance to colic
19	36	75.0	597	1 HQECL	hydrogenase (EC 1.
20	36	75.0	597	2 AE0721	hydrogenase-1 larg
21	36	75.0	597	2 E85632	hydrogenase-1 larg
22	36	75.0	597	2 A90770	hydrogenase-1 larg
23	36	75.0	1711	1 A47392	chromodomain-helic
24	35	72.9	77	2 B82179	probable glutaredo
25	35	72.9	246	2 C83694	hypothetical prote
26	35	72.9	363	2 T26842	hypothetical prote
27	35	72.9	499	2 AC0882	probable amino aci
28	35	72.9	565	2 T47423	hypothetical prote
29	35	72.9	566	2 B82173	probable ABC trans

30	35	72.9	1435	2 D96693	protein Putative A
31	35	72.9	2228	1 ZLNZSV	genome polypoteine
32	35	72.9	3149	1 Q0BE8	Bp1 protein - hu
33	34	70.8	59	2 A43985	myosin-light-chain
34	34	70.8	207	2 JH0145	hypothetical 24.1K
35	34	70.8	228	2 A31403	membrane protein B
36	34	70.8	272	2 J04170	trypsin-like prote
37	34	70.8	288	2 T10477	sec3 protein - ye
38	34	70.8	291	2 S61498	chemotactic methyl
39	34	70.8	297	2 T39905	protein transport
40	34	70.8	297	2 A45442	transport vesicle
41	34	70.8	336	2 A32507	41K larval antigen
42	34	70.8	360	2 T18140	hypothetical prote
43	34	70.8	360	2 B71073	probable malate de
44	34	70.8	381	2 H81985	alpha-D-mannose-al
45	34	70.8	395	2 S25851	calreticulin precu

ALIGNMENTS

```
RESULT 1
E90973
hypothetical protein ECs2757 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E90973
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G
gaawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90973
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <HAV>
A:Cross-references: UNIPROT:O8X4V0; GB:BA000007; PIDD:BA036180.1; PID:g1336225; GSPDB:(
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECs2757

Query Match      81.2% Score 39; DB 2; Length 103;
Best Local Similarity 50.0%; Pred. No. 4.8;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY      1 WXPQFER 8
Db      17 WTHPEYK 24

RESULT 2
unknown protein encoded within prophage CP-931U [imported] - Escherichia coli (strain O1
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: A85821
R:Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobbeck, E.J.; Davis, N.W.; Lam, A.; Dimmlante, E.; Potamouists, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85821
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <STO>
A:Cross-references: UNIPROT:O8X4V0; GB:AE005174; NID:g12516136; PIDD:AA057029.1; GSPDB:(
A:Experimental source: strain O157:H7, substrain ED2933
C:Genetics:
A:Gene: Z3120

Query Match      81.2% Score 39; DB 2; Length 103;
Best Local Similarity 50.0%; Pred. No. 4.8;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
```

OY 1 WKHPQFER 8
Db 17 WTHPEYER 24

RESULT 3

C85713
Unknown protein encoded within prophage CP-9330 [imported] - Escherichia coli (strain O1
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C/Accession: C85713
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A/Reference number: A85480; MUID:21074935; PMID:11206551
A/Accession: C85713
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-103 <STO>
A/Cross-references: UNIPROT:Q8X549; GB:AE005174; NID:g12515046; PIDN:AA56167.1; GSPDB:C
A/Experimental source: strain O157:H7, substrain EDL533
C/Genetics:
A/Gene: Z2097

Query Match 81.2%; Score 39; DB 2; Length 103;
Best Local Similarity 50.0%; Pred. No. 4.8;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 WKHPQFER 8
Db 17 WTHPEYER 24

RESULT 4

A90904
Hypothetical protein Ec82201 [imported] - Escherichia coli (strain O157:H7, substrain RI
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C/Accession: A90904
R/Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Gasaawa, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A/Reference number: A99629; MUID:21156231; PMID:11258796
A/Accession: A90904
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-103 <HAV>
A/Cross-references: UNIPROT:Q8X549; GB:BA000007; PIDN:BA35624.1; PID:g13361667; GSPDB:C
A/Experimental source: strain O157:H7, substrain RMD 0509552
C/Genetics:
A/Gene: Ec82201

Query Match 81.2%; Score 39; DB 2; Length 103;
Best Local Similarity 50.0%; Pred. No. 4.8;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 WKHPQFER 8
Db 17 WTHPEYER 24

RESULT 5

PC4427
Nck, Ash and phospholipase C gamma-binding protein 4 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 04-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 09-Jul-2004
C/Accession: PC4427
R/Matucka, K.; Miki, H.; Takahashi, K.; Takenawa, T.
Biochem. Biophys. Res. Commun. 239, 488-492, 1997
A/Title: A novel ligand for an SH3 domain of the adaptor protein Nck bears an SH2 domain

A/Reference number: PC4427; MUID:98008866; PMID:9344857

A/Accession: PC4427

A/Status: nucleic acid sequence not shown

A/Molecule type: mRNA

A/Residues: 1-485 <MAT>
A/Cross-references: UNIPROT:O14512; DBJ:AB005216; NID:g2443366; PIDN:BAA22432.1; PID:g-
F/1-43/65-78/180-191/220-230/Region: proline-rich
F/277-414/Domain: SH2 homology <SH2>

Query Match 81.2%; Score 39; DB 2; Length 485;
Best Local Similarity 71.4%; Pred. No. 26;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 WKHPQFE 7
Db 336 WCHPKE 342

RESULT 6

T20278
Hypothetical protein ZK256.1b - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Aug-2000
C/Accession: T20278; T27797
R/Kershaw, J.
Submitted to the EMBL Data Library, November 1996
A/Reference number: Z19246
A/Accession: T20278
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-705 <WIL>
A/Cross-references: EMBL:Z81490; PIDN:CA804016.1; GSPDB:GN00019; CESP:ZK256.1b
A/Experimental source: clone CC4
R/McLay, K.
Submitted to the EMBL Data Library, November 1996
A/Reference number: Z20420
A/Accession: T27797
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-705 <WIL>
A/Cross-references: EMBL:Z82088; PIDN:CA805001.1; GSPDB:GN00019; CESP:ZK256.1b
A/Experimental source: clone ZK256
C/Genetics:
A/Gene: CESP:ZK256.1b
A/Map position: 1
A/Intons: 104/3; 181/3; 273/3; 465/1; 566/1; 661/3; 694/2
A/Intons: 104/3; 181/3; 273/3; 465/1; 566/1; 661/3; 694/2
A/Superfamily: H+/K+-transporting ATPase chain B; ATPase nucleotide-binding domain homol.
F/525-695/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match 81.2%; Score 39; DB 2; Length 705;
Best Local Similarity 62.5%; Pred. No. 40;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 WKHPQFER 8
Db 383 WTHPEFAR 390

RESULT 7

T20277
Hypothetical protein ZK256.1a - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C/Accession: T20277; T27796
R/Kershaw, J.
Submitted to the EMBL Data Library, November 1996
A/Reference number: Z19246
A/Accession: T20277
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-922 <WIL>

Wed Mar 2 14:27:08 2005

A:Cross-references: UNIPROT:Q9XTG4; EMBL:Z81490; PIDN:CAB04015.1; GSPDB:GN00019; CESP:ZK
A:Experimental source: clone CCA
R:McLay, K.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z20420
A:Accession: T27796
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-922 <MT2>
A:Cross-references: EMBL:Z82088; PIDN:CAB05000.1; GSPDB:GN00019; CESP:ZK256.1a
A:Experimental source: clone ZK256
C:Gene: CESP:ZK256.1a
C:Genetics:
A:Map position: 1
A:Intron: 104/3; 181/3; 273/3; 465/1; 566/1; 661/3; 694/2; 782/3; 884/1
C:Superfamily: Na+/K+-transporting ATPase alpha chain; ATPase nucleotide-binding domain
F:1525-695/Domain: ATPase nucleotide-binding domain homology <ATN>

Query Match 81.2%; Score 39; DB 2; Length 922;
Best Local Similarity 62.5%; Pred. No. 53;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXPQPER 8
DB 383 WTHPEFAR 390

RESULT 8
S66537
Ig heavy chain V region - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: S66537
R:Tsichlis, G.; Haase, W.; Engel, A.; Michel, H.
Eur. J. Biochem. 231, 823-830, 1995
A:Title: Isolation and structural characterization of trimeric cyanobacterial photosystem
A:Reference number: S66536; MUID:95377318; PMID:7649183
A:Accession: S66537
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-131 <TS1>
A:Cross-references: EMBL:X88902; NID:9895869; PIDN:CAA61364.1; PID:g1103701
C:Superfamily: immunoglobulin V region; immunoglobulin homology
F:15-98/Domain: immunoglobulin homology <IMM>

Query Match 79.2%; Score 38; DB 2; Length 131;
Best Local Similarity 83.3%; Pred. No. 9.4;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXPQPER 6
DB 124 WXPQPER 129

RESULT 9
XJRPFK
transketolase (EC 2.2.1.1) - Rhodobacter sphaeroides
N:Alternate names: glycolaldehydetransferase
C:Species: Rhodobacter sphaeroides
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
C:Accession: B41080
R:Chen, J.H.; Gibson, J.L.; McGue, L.A.; Tdbite, F.R.
J. Biol. Chem. 266, 20447-20452, 1991
A:Title: Identification, expression, and deduced primary structure of transketolase and
A:Reference number: A41080; MUID:92041881; PMID:1939098
A:Accession: B41080
A:Molecule type: DNA
A:Residues: 1-657 <CHS>
A:Cross-references: UNIPROT:P29277; GB:M68914; NID:g151588; PIDN:AAA26155.1; PID:g151590
A>Note: the authors translated the codon GGC for residue 173 as Ala
C:Comment: The active enzyme catalyzes the transfer of a keto group to an aldehyde acceptor
C:Superfamily: transketolase; thiamin pyrophosphate-binding domain homology

C:Keywords: Calvin cycle; homodimer; magnesium; pentose phosphate pathway; thiamin pyrophosphate-binding domain homology <TPB>
F:147-197/Domain: thiamin pyrophosphate-binding domain homology <TPB>

Query Match 77.1%; Score 37; DB 1; Length 657;
Best Local Similarity 71.4%; Pred. No. 84;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXPQPER 7
DB 280 WDPPEPE 286

RESULT 10
AP0815
transketolase (EC 2.2.1.1) - Salmonella enterica subsp. enterica serovar Typhi (strain C)
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A>Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AP0815
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AP0815
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-666 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07704.1; PID:g16503690; GSPDB:GN00176
C:Genetics:
A:Gene: STY2711
C:Superfamily: transketolase; thiamin pyrophosphate-binding domain homology
C:Keywords: transferase

Query Match 77.1%; Score 37; DB 2; Length 666;
Best Local Similarity 71.4%; Pred. No. 85;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXPQPER 7
DB 278 WHPPEPE 284

RESULT 11
A48660
transketolase (EC 2.2.1.1) B - Escherichia coli (strain K-12)
N:Alternate names: glycolaldehydetransferase B; transketolase (EC 2.2.1.1) 2
C:Species: Escherichia coli
C:Date: 16-Feb-1994 #sequence_revision 04-Oct-1996 #text_change 09-Jul-2004
C:Accession: A48660; H65021
R:Iida, A.; Teshiba, S.; Mizobuchi, K.
J. Bacteriol. 175, 5375-5383, 1993
A:Title: Identification and characterization of the tkpB gene encoding a second transketolase
A:Reference number: A48660; MUID:93374831; PMID:8396116
A:Accession: A48660
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-667 <IID>
A:Cross-references: UNIPROT:P33570; GB:D12473; NID:g440349; PIDN:BA02039.1; PID:g460975
A:Experimental source: strain K-12
A>Note: sequence extracted from NCBI backbone (NCBIN:137560, NCBI:137561)
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A67720; MUID:97426617; PMID:9278503
A:Accession: H65021
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-667 <BLAT>
A:Cross-references: GB:AE000333; GB:U00096; NID:g1788805; PIDN:AACT5518.1; PID:g1788808;

A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: tkbB
A:Map position: 53 min
C:Complex: homodimer
C:Function:
A:Description: catalyzes the formation of D-ribose 5-phosphate and D-xylose 5-phosphat
C:Superfamily: transketolase; thiamin pyrophosphate-binding domain homology
C:Keywords: Calvin cycle; homodimer; magnesium; pentose phosphate pathway; thiamin pyrog
F:143-193/Domain: thiamin pyrophosphate-binding domain homology <RPB>

Query Match
Best Local Similarity 77.1%; Score 37; DB 1; Length 667;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WXPQFE 7
Db 278 WHHPFE 284

RESULT 12
G91044
transketolase 2 isozyme [imported] - Escherichia coli (strain O157:H7, substrain RIMD 05
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: G91044
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gaawara, N.; Yasunaga, T.; Kohara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Rep. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; MUID:2115631; PMID:11258796
A:Accession: G91044
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-667 <RAY>
A:Cross-references: UNIPROT:Q8XBFI; GB:BA000007; PIDN:BAB36750.1; PID:g13362797; GSPDB:C
C:Genetics:
A:Gene: ECG327
C:Superfamily: transketolase; thiamin pyrophosphate-binding domain homology

Query Match
Best Local Similarity 77.1%; Score 37; DB 2; Length 667;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WXPQFE 7
Db 278 WHHPFE 284

RESULT 13
B85889
transketolase 2 isozyme [imported] - Escherichia coli (strain O157:H7, substrain EDL933)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: B85889
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Llim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85889
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-667 <STO>
A:Cross-references: UNIPROT:Q8XBFI; GB:AE005174; NID:g12516840; PIDN:ANG57574.1; GSPDB:C
C:Genetics:
A:Gene: tkbB
C:Superfamily: transketolase; thiamin pyrophosphate-binding domain homology

Query Match
Best Local Similarity 77.1%; Score 37; DB 2; Length 667;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WXPQFE 7
Db 278 WHHPFE 284

RESULT 14
H69297
conserved hypothetical protein AF0384 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
C:Accession: H69297
R:Kleink, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kyrleese, E.F.
Glodek, A.; Zhou, L.; Overbeek, R.; Goeyne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Overbeek, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Woese, C.R.; Venter, J.C.
A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeac
A:Reference number: A69250; MUID:98049343; PMID:9389475
A:Accession: H69297
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-313 <KLB>
A:Cross-references: UNIPROT:Q29863; GB:AE001078; GB:AE000782; NID:g2689401; PIDN:AAB9085

Query Match
Best Local Similarity 75.0%; Score 36; DB 2; Length 313;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPQFE 8
Db 266 WYHFEFR 273

RESULT 15
H83444
probable cytochrome c PA1600 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: H83444
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.D.; Br
adman, S.; Yun, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kass, A.; Larbig, K.; Llim,
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A83950; MUID:20437337; PMID:10984043
A:Accession: H83444
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-433 <STO>
A:Cross-references: UNIPROT:Q913C1; GB:AE004588; GB:AE004091; NID:g9947563; PIDN:ANG0498
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1600
C:Superfamily: membrane-bound alcohol dehydrogenase, cytochrome c subunit; cytochrome c
C:Keywords: chromoprotein; heme; iron; metalloprotein
F:62/Binding site: heme (Cys) (covalent) #status predicted
F:62/Binding site: heme iron (His) (covalent) #status predicted
F:204/207/Binding site: heme (Cys) (covalent) #status predicted
F:208/Binding site: heme iron (His) (axial ligand) #status predicted
F:336/339/Binding site: heme (Cys) (covalent) #status predicted
F:340/Binding site: heme iron (His) (axial ligand) #status predicted

Query Match
Best Local Similarity 75.0%; Score 36; DB 2; Length 433;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WXPQFE 8
Db 99 WXPQFE 106

Wed Mar 2 14:27:08 2005

seq3.rpr

Page 5

Search completed: March 2, 2005, 12:28:49
Job time : 7.11382 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 29.7236 Seconds
(without alignments)
137.824 Million cell updates/sec

Title: SEQ3
Perfect score: 48
Sequence: 1 wxhpqfer 8

Scoring table: BLOSUM62DX
Gapop 10.0, Gapext 0.5

Searched: 1612378 seqs, 512079187 residues
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: Uniprot.03:*
1: uniprot_sprot:*
2: uniprot_crembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query length	DB ID	Description
1	45	93.8	704	2	08CJD2
2	45	93.8	704	2	08CH90
3	40	83.3	130	2	07GS79
4	40	83.3	295	2	063KN1
5	40	83.3	296	2	062CE8
6	40	83.3	312	2	09V6M2
7	40	83.3	315	2	08A7X7
8	40	83.3	429	2	09F6L2
9	40	83.3	508	1	MATK_MARRSC
10	40	83.3	508	2	09B8G0
11	40	83.3	508	2	09B8G3
12	39	81.2	103	2	08X549
13	39	81.2	103	2	08X4V0
14	39	81.2	485	1	SOC7_HUMAN
15	39	81.2	532	1	AD50_BOVIN
16	39	81.2	579	1	SOC7_MOUSE
17	39	81.2	633	3	07UYJ7
18	39	81.2	705	2	06LA78
19	39	81.2	901	2	09XTG4
20	39	81.2	978	2	06LA80
21	39	81.2	1335	2	09FNP5
22	38	79.2	127	2	065QC7
23	38	79.2	176	2	07ID58
24	38	79.2	176	2	09VXD1
25	38	79.2	202	2	06E071
26	38	79.2	251	2	09FZU5
27	38	79.2	263	2	062G83
28	38	79.2	277	2	08S0G7
29	38	79.2	295	2	09APW1
30	38	79.2	295	2	088J42
31	38	79.2	620	2	082CV1

32	38	79.2	634	2	093J80	Q93180 streptomyce
33	38	79.2	1915	2	07Y1B7	Q7Y1B7 hordem vul
34	37	77.1	375	2	06D7N4	Q6D7N4 erwina car
35	37	77.1	376	2	07X930	Q7X930 prunus avlu
36	37	77.1	384	2	07VS20	Q7VS20 bordeteila
37	37	77.1	384	2	07WEC2	Q7WEC2 bordeteila
38	37	77.1	455	2	06MKM3	Q6MKM3 bdellovibri
39	37	77.1	459	2	067PY8	Q67PY8 symbiobacte
40	37	77.1	508	2	09B8G8	Q9B8G8 apinagia su
41	37	77.1	557	2	08XRVO	Q8XRVO ralestonia s
42	37	77.1	594	2	0750N1	Q750N1 ashbya gos
43	37	77.1	643	2	08TXV8	Q8TXV8 methanopyru
44	37	77.1	655	2	06N0W7	Q6N0W7 rhodospseudo
45	37	77.1	657	1	TXT_RHOSH	P29277 rhodobacter

ALIGNMENTS

RESULT 1	
08CJD2	PRELIMINARY; PRT; 704 AA.
AC 08CJD2;	
DT 01-MAR-2003 (TREMBLrel. 23, Created)	
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)	
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)	
DE Guanylyl cyclase alpha 1 subunit.	
OS Rattus norvegicus (Rat).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX NCBI_Taxid=10116;	
RN [1]	
RP SEQUENCE FROM N.A.	
RA Nakamura I., Suzuki N.;	
RL Submitted (NOV-2002) to the EMBL/Genbank/DBJ databases.	
DR EMBL; AB096020; BAC24016.1; -	
DR HSP; P30803; IAZS.	
DR GO; GO:0004383; F:guanylate cyclase activity; IEA.	
DR GO; GO:0016829; F:lyase activity; IEA.	
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.	
DR InterPro; IPR01054; G_cyclase.	
DR Pfam; PF00211; Guanylate_cyc; 1.	
DR SMART; SM0044; CYC; 1.	
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.	
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.	
KW Lyase.	
SQ SEQUENCE 704 AA; 79188 MW; F73FC97B68544E6A CRC64;	
Query Match	93.8%; Score 45; DB 2; Length 704;
Best Local Similarity	75.0%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;	
Qy 1 WXHPQFER 8	
Db 697 WSHPOPEK 704	
RESULT 2	
08CH90	PRELIMINARY; PRT; 743 AA.
AC 08CH90;	
DT 01-MAR-2003 (TREMBLrel. 23, Created)	
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)	
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)	
DE Soluble guanylyl cyclase alpha 2 subunit E219G mutant.	
OS Rattus norvegicus (Rat).	
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
OX NCBI_Taxid=10116;	
RN [1]	
RP SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Nakamura I., Yao Y., Suzuki N.;	

```

RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB097860; BAC44887.1; -.
DR HSBP; P30803; IAZS.
DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0007242; P:intracellular signalling cascade; IEA.
DR InterPro; IPR001054; G: cyclase.
DR InterPro; IPR009080; rRNAyn_1a_bind.
DR Pfam; PF00211; Guanylate_cyc; 1.
DR SMART; SM00044; CYC; 1.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
DR KW Lyase.
SQ SEQUENCE 743 AA; 83251 MW; ACF5C53E092813A CRC64;

Query Match
Best Local Similarity 93.3%; Score 45; DB 2; Length 743;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFER 8
Db 736 WSHPOFER 743

RESULT 3
Q70S79 PRELIMINARY; PRT; 130 AA.
ID Q70S79;
AC Q70S79;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE GIP_149_16288_16680.
OS Giardia lamblia ATCC 50803.
OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
ON NCBI_TaxID=184922;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WB C6;
RA Morrison H.G., McArthur A.G., Adam R.D., Aley S.B., Gillin F.D.,
RT Olsen G.J., Sogin M.L.;
RT "Draft sequence of the Giardia lamblia genome.";
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACB01000123; BAA37879.1; -.
SQ SEQUENCE 130 AA; 14514 MW; FAC704BD76FD0B8A CRC64;

Query Match
Best Local Similarity 83.3%; Score 40; DB 2; Length 130;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXHPOFER 8
Db 7 WLBHYER 14

RESULT 4
Q63KN1 PRELIMINARY; PRT; 295 AA.
ID Q63KN1;
AC Q63KN1;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Hypothetical protein.
GN ORNames=BPS1330;
OS Burkholderia pseudomallei K96243.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=272560;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K96243;

```

```

RX PubMed=15377794;
RA Holden M.T.G., Tibball R.W., Peacock S.U., Cerdano-Tarraga A.M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.,
RA Bentley S.D., Sebahia M., Thomson N.R., Bacon N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Chevreton I.,
RA Chillingworth T., Cronin A., Crosset B., Davis P., Deshaizer D.,
RA Felwell T., Fraser A., Hance Z., Hauser H., Holroyd S., Jagels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinowitch E., Rutherford K., Sanders M., Simmonds M.,
RA Songstival S., Stevens K., Tumapa S., Vesatchavest M.,
RA Whitehead S., Yeats C., Barrett B.G., Oyston P.C.P., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).
DR EMBL; BX571966; CAH38801.1; -.
SQ SEQUENCE 295 AA; 32602 MW; 87CD7AC960FAA972 CRC64;

Query Match
Best Local Similarity 83.3%; Score 40; DB 2; Length 295;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXBPOFER 8
Db 247 WPHPOFER 254

RESULT 5
Q62CE8 PRELIMINARY; PRT; 296 AA.
ID Q62CE8;
AC Q62CE8;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)
DE Hydrolyase.
GN ORNames=BMAA0947;
OS Burkholderia mallei ATCC 23344.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=243160;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 23344;
RA Nieman W.C.; Deshaizer D., Kim H.S., Tettelin H., Nelson K.E.,
RA Feldblyum T., Ulrich R.L., Romning C.M., Birkac L.M., Daugherty S.C.,
RA Davidson T.D., Deboy R.T., Dimitrov G., Dodson R.J., Durkin A.S.,
RA Gwyn M.L., Haft D.H., Khouri H., Kolonay J.F., Madupu R.,
RA Mohammad Y., Nelson W.C., Radune D., Romero C.M., Sarría S.,
RA Selengut J., Shamlin C., Sullivan S.A., White O., Yu Y., Zafar N.,
RA Zhou L., Fraser C.M.;
RT "Structural flexibility in the Burkholderia mallei genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14247-14251(2004).
DR EMBL; CP000011; AAU46133.1; -.
KW Hydrolyase.
SQ SEQUENCE 296 AA; 32617 MW; DA8DBB20B3581A4 CRC64;

Query Match
Best Local Similarity 83.3%; Score 40; DB 2; Length 296;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WXBPOFER 8
Db 248 WPHPOFER 255

RESULT 6
Q9V6M2 PRELIMINARY; PRT; 312 AA.
ID Q9V6M2;
AC Q9V6M2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE CG12765-PA.

```

```

GN ORFNames=CG12765;
OS Drosophila melanogaster (Fruit fly)
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Change M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abitl J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brockett P., Brottier P.,
RA Burdick K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hoselt D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milbina N.V., Modary C., Morris J., Mostreli A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paciel J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheele F., Shen H.,
RA Shue B.C., Siders-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirekas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodruff W., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zdevri J.S., Zhan M., Zhang G., Zhao X., Zhu L.,
RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Ceiniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Change M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Paciel J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirekas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirekas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Ceiniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomic perspective."
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Miera S., Crocby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,

```

```

RA Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Ceiniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN [5]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RX FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB003819; AAF58401.2; -
DR FlyBase; Fgmn033813; CG12765.
DR InterPro; IPR001810; F-box.
DR Pfam; PF00646; F-box; 1.
SQ SEQUENCE 312 AA; 36584 MW; BB81706977A69B03 CRC64;

Query Match
Best Local Similarity 83.3%; Score 40; DB 2; Length 312;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 WKHPQPER 8
DB 302 WHPDFOR 309

RESULT 7
Q8A7X7 PRELIMINARY; PRT; 315 AA.
ID Q8A7X7;
AC Q8A7X7;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative membrane protein.
GN Ordered locus names=BT1397;
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=2250858; PubMed=1263928; DOI=10.1126/science.1080029;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis."
RL Science 299:2074-2076 (2003).
DR EMBL; AB016931; AAC76504.1; -
GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR00620; DUF6.
DR Pfam; PF00892; DUF6; 2.
KW Complete proteome.
SQ SEQUENCE 315 AA; 34785 MW; 08782D50EB6D83C3 CRC64;

Query Match
Best Local Similarity 83.3%; Score 40; DB 2; Length 315;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 WKHPQPER 8
DB 154 WHPQSER 161

RESULT 8
Q9F6L2 PRELIMINARY; PRT; 429 AA.
ID Q9F6L2;
AC Q9F6L2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)

```

```

DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
GN SgrAIM methylase.
OS Streptomyces griseus.
OC Bacteria; Actinobacteria; Actinomycetales;
OX Streptomycinae; Streptomycetaceae; Streptomyces.
RN NCB1_TaxID=1911;
RP SEQUENCE FROM N.A.
RA Higgins L.S., Kong H.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + DNA = S-adenosyl-L-
homocysteine + DNA containing 5-methylcytosine.
CC -1- SIMILARITY: Belongs to the C5-methyltransferase family.
DR EMBL; AF290880; AAC31558.1; -.
DR HSBP; P20589; IDCT.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0008168; P:methyltransferase activity; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0006306; P:DNA methylation; IEA.
DR GO; GO:0009307; P:DNA restriction; IEA.
DR InterPro; IPR001525; C5_DNA_meth.
DR Pfam; PF00145; DNA_methylase; 1.
DR PRINTS; PR00105; CSMETTRFASR.
DR TIGRFAMs; TIGR00675; dcm; 1.
DR PROSITE; PS00094; C5_MTASE_1; 1.
KM Methyltransferase; Restriction system; Transferase.
SQ SEQUENCE 429 AA; 48321 MW; 3E78F7859D088120 CRC64;

Query Match 83.3%; Score 40; DB 2; Length 429;
Best Local Similarity 62.5%; Pred. No. 81;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 WKXPOFER 8
DB 348 WHPDYER 355

RESULT 9
MATK_MARSC STANDARD; PRT; 508 AA.
ID MATK_MARSC
AC O8WK9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Maturase K (intron maturase).
GN Name=matk;
OS Marathrum schiedeanum.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eustoids I; Malpighiales; Podostemaceae; Marathrum.
OX NCB1_TaxID=116737;
RN [1]
RP SEQUENCE FROM N.A.
RA Kita Y., Kato M.;
RL "Phylogenetic relationships of the aquatic angiosperm family
Podostemaceae inferred from matk sequence data."
Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Probably assists in splicing chloroplast group II
introns (By similarity).
CC -1- SIMILARITY: Belongs to the intron maturase family 2. Matk
subfamily.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----

```

```

DR EMBL; AB038195; BAB83156.1; -.
DR InterPro; IPR008998; Agglutinin.
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; Matk_N.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; Matk_N; 1.
KM Chloroplast; mRNA processing.
SQ SEQUENCE 508 AA; 61342 MW; ECCF5B416B0AB5C3 CRC64;

Query Match 83.3%; Score 40; DB 1; Length 508;
Best Local Similarity 62.5%; Pred. No. 97;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 WKXPOFER 8
DB 188 WHPDYER 195

RESULT 10
Q9BGO PRELIMINARY; PRT; 508 AA.
ID Q9BGO
AC Q9BGO;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Maturase K.
GN Name=matk;
OS Vanoyenella plumosa.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eustoids I; Malpighiales; Podostemaceae; Vanoyenella.
OX NCB1_TaxID=51609;
RN [1]
RP SEQUENCE FROM N.A.
RA Kita Y., Kato M.;
RL "Infrafamilial Phylogeny of the Aquatic Angiosperm Podostemaceae
inferred from the Nucleotide Sequences of the matk Gene."
Plant Biol. 3:156-163(2001).
DR EMBL; AB048378; BAB33398.1; -.
DR GO; GO:0009507; C:chloroplast; IEA.
DR GO; GO:0008380; P:RNA splicing; IEA.
DR InterPro; IPR006998; Agglutinin.
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; Matk_N.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; Matk_N; 1.
KM Chloroplast.
SQ SEQUENCE 508 AA; 61288 MW; 90229C45138AF4B6 CRC64;

Query Match 83.3%; Score 40; DB 2; Length 508;
Best Local Similarity 62.5%; Pred. No. 97;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 WKXPOFER 8
DB 188 WHPDYER 195

RESULT 11
Q9BGO PRELIMINARY; PRT; 508 AA.
ID Q9BGO
AC Q9BGO;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Maturase K.
GN Name=matk;
OS Oserya coulteriana.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eustoids I; Malpighiales; Podostemaceae; Oserya.

```

```

OX NCBI_TaxID=51602;
RN [1]
RP SEQUENCE FROM N.A.
RA Kita Y., Kato M.,
RT "Infrafamilial Phylogeny of the Aquatic Angiosperm Podostemaceae
RT Inferred from the Nucleotide Sequences of the matK Gene.",
RL Plant Biol. 3:156-163(2001).
DR EMBL; AB048375; BAB33395.1; -.
DR GO; GO:0009507; Chloroplast; IEA.
DR GO; GO:0008380; rRNA splicing; IEA.
DR InterPro; IPR008998; Agglutinin.
DR InterPro; IPR000442; Intron mature2.
DR Pfam; PF01348; Intron mature2; 1.
DR Pfam; PF01824; MatK_N; 1.
KM Chloroplast.
SQ SEQUENCE 508 AA; 61400 MW; 45240C7D31B54CDF CRC64;

Query Match      83.3%; Score 40; DB 2; Length 508;
Best Local Similarity 62.5%; Pred. No. 97;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 WXPQFER 8
Db 188 WSHPSFEK 195

RESULT 12
ID Q8X549 PRELIMINARY; PRT; 103 AA.
AC Q8X549; Q7ADV0;
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein Z2097 (Hypothetical protein Ec2201).
GN OrderedLocName=EC2201, Z2097;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AF005346; AAG56167.1; -.
DR EMBL; AF002557; BAB35624.1; -.
DR PIR; A90904; A90904.
DR PIR; C85713; C85713.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 103 AA; 11980 MW; 5A47A9F91CF29EB6 CRC64;

Query Match      81.2%; Score 39; DB 2; Length 103;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 WXPQFER 8
Db 17 WTHPEYEK 24

RESULT 13
ID Q8X4V0 PRELIMINARY; PRT; 103 AA.
AC Q8X4V0; Q7ACU2;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein Z3120 (Hypothetical protein Ec2757).
GN OrderedLocName=Ec2757, Z3120;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952 / EHEC;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AF005423; AAG57029.1; -.
DR EMBL; AP002559; BAB36180.1; -.
DR PIR; A85821; A85821.
DR PIR; E90973; E90973.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 103 AA; 12092 MW; D4F3CEA5A1089D30 CRC64;

Query Match      81.2%; Score 39; DB 2; Length 103;
Best Local Similarity 50.0%; Pred. No. 26;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXPQFER 8
Db 17 WTHPEYEK 24

RESULT 14
ID SOC7 HUMAN STANDARD; PRT; 485 AA.
AC O14512;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 25-OCT-2004 (Rel. 45, Last annotation update)
DE Suppressor of cytokine signalling 7 (SOCS-6) (Nck, Ash and
DE phospholipase C gamma-binding protein) (Nck-associated protein 4)
DE (NAP-4) (Fragment).
GN Name=SOCS7; Synonym=NAP4, SOCS6;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;

```

```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=98008966; PubMed=9344857; DOI=10.1006/dbrc.1997.7492;
RA      Matlock K., Miki H., Takahashi K., Takenawa T.;
RT      "A novel ligand for an SH3 domain of the adaptor protein Nck bears an
RL      SH2 domain and nuclear signaling motifs."
RL      Biochem. Biophys. Res. Commun. 239:488-492(1997).
CC      -1- FUNCTION: SOCS family proteins form part of a classical negative
CC      feedback system that regulates cytokine signal transduction.
CC      -1- SUBUNIT: Interacts, via the third proline-rich region, with the
CC      second SH3 domain of the adapter protein NCK. Also interacts with
CC      GRB2 and phospholipase C-gamma.
CC      -1- TISSUE SPECIFICITY: Expressed in brain and leukocytes. Also in
CC      fetal lung fibroblasts and fetal brain.
CC      -1- SIMILARITY: Contains 1 SH2 domain.
CC      -1- SIMILARITY: Contains 1 SOCS box domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AB005216; BAA22432.1; -.
CC      FIRM; PC4427; PC4427.
CC      HSSP; P23727; 1004.
CC      DR      Genew; HGNC; 29846; SOCS7.
CC      MIM; 608788; -.
CC      DR      GO; GO:0017124; F:SH3-domain binding; NAS.
CC      DR      InterPro; IPR000980; SH2.
CC      DR      InterPro; IPR001496; SOCS_C.
CC      DR      Pfam; PF07525; C1ip; 1.
CC      DR      Pfam; PF00017; SH2; 1.
CC      DR      ProDom; PD000093; SH2; 1.
CC      DR      SMART; SM00262; SH2; 1.
CC      DR      SMART; SM00253; SOCS; 1.
CC      DR      PROSITE; PSS0001; SH2; 1.
CC      DR      PROSITE; PSS0225; SOCS; 1.
CC      KM      Growth regulation; SH2 domain; signal transduction inhibitor.
CC      FT      NON_TER      1
CC      FT      DOMAIN      1
CC      FT      DOMAIN      31
CC      FT      DOMAIN      180
CC      FT      DOMAIN      260
CC      FT      DOMAIN      277
CC      FT      DOMAIN      381
CC      FT      DOMAIN      431
CC      FT      DOMAIN      18
CC      FT      DOMAIN      25
CC      FT      DOMAIN      57
CC      FT      DOMAIN      64
CC      SQ      SEQUENCE 485 AA; 53564 MW; 0CC5EC107174A4F9 CRC64;

Query Match      81.2%; Score 39; DB 1; Length 485;
Best Local Similarity 71.4%; Pred. NO. 1.4e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 WXPQPER 7
DB      336 WCHPKFE 342

RESULT 15
ADSO_BOVIN
ID      ADSO_BOVIN      STANDARD; PRT; 532 AA.
AC      Q27969;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      05-JUL-2004 (Rel. 44, Last annotation update)
DE      Adrenal medulla 50 kDa protein.
GN      Name=ADP50;
OS      Bos taurus (Bovine).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC      Bovinae; Bos.

```

```

OX      NCBI_TaxID=9913;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Adrenal medulla;
RA      Nelson N., Nelson H., Mandiyan S.;
RT      Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
RL      Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; U04706; AAA17986.1; -.
CC      DR      InterPro; IPR000413; Integrin_alpha.
CC      DR      Pfam; PF01839; FG-GAP; 2.
CC      SQ      SEQUENCE 532 AA; 58160 MW; B0B1BE87194369B CRC64;

Query Match      81.2%; Score 39; DB 1; Length 532;
Best Local Similarity 75.0%; Pred. NO. 1.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 WXPQPER 8
DB      247 WXPQPER 254

Search completed: March 2, 2005, 12:44:11
Job time : 30.7236 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 36.0325 Seconds
(without alignments)
85.869 Million cell updates/sec

Title: SEQ3

Perfect score: 48

Sequence: 1 wxhpgfer 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 366760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq.16Dec04:*

1: geneseqp1980s:*\n2: geneseqp1990s:*\n3: geneseqp2000s:*\n4: geneseqp2001s:*\n5: geneseqp2002s:*\n6: geneseqp2003as:*\n7: geneseqp2003bs:*\n8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	11	2	AAR52694
2	48	100.0	11	2	AAR52693
3	48	100.0	12	2	AAR52697
4	45	93.8	8	2	AAR59212
5	45	93.8	8	4	AAB35433
6	45	93.8	8	4	AAB35442
7	45	93.8	8	4	AAB68616
8	45	93.8	8	5	AA019065
9	45	93.8	8	5	ABBS7464
10	45	93.8	8	5	ABB77486
11	45	93.8	8	6	ABG73584
12	45	93.8	8	6	AAE37230
13	45	93.8	8	6	ABP60361
14	45	93.8	8	6	ABP60368
15	45	93.8	8	7	ADA09808
16	45	93.8	8	7	ADB84588
17	45	93.8	8	7	ADB85500
18	45	93.8	8	7	ADD29930
19	45	93.8	8	8	AD006951
20	45	93.8	8	8	ADP90832
21	45	93.8	8	8	ADG20244
22	45	93.8	8	8	ADR72509
23	45	93.8	9	5	ABG31054
24	45	93.8	10	4	AAV97662
25	45	93.8	10	4	AAB99027

26	45	93.8	10	5	AAU80475	Peptide S
27	45	93.8	10	6	ABP56623	Abp56623 C-terminus
28	45	93.8	10	8	ADP11064	Adp11064 Strep-tag
29	45	93.8	10	8	ADN16967	Adn16967 Human res
30	45	93.8	10	8	ADO26489	Ado26489 Strep-tag
31	45	93.8	11	6	AAE38373	Aae38373 Epitope t
32	45	93.8	11	8	ADG20228	Adg20228 Strep tag
33	45	93.8	19	6	ABG74881	Abg74881 Bacteriop
34	45	93.8	19	6	ABG74882	Abg74882 Bacteriop
35	45	93.8	19	6	ABG74880	Abg74880 Bacteriop
36	45	93.8	19	8	AD139157	Ad139157 Construct
37	45	93.8	19	8	AD139155	Ad139155 Construct
38	45	93.8	19	8	AD139156	Ad139156 Construct
39	45	93.8	24	6	ABP60363	Abp60363 Strep-tag
40	45	93.8	24	6	ABP60362	Abp60362 Strep-tag
41	45	93.8	35	6	ABP60369	Abp60369 Strep-tag
42	45	93.8	36	6	ABP60370	Abp60370 Strep-tag
43	45	93.8	117	5	AAU97558	AAU97558 Synthetic
44	45	93.8	117	5	AAU97553	AAU97553 Synthetic
45	45	93.8	117	5	AAU97557	AAU97557 Synthetic

ALIGNMENTS

RESULT 1
AAR52694 standard; protein; 11 AA.
ID AAR52694
XX AAR52694;
AC 10-JAN-1995 (first entry)
XX
DT 10-JAN-1995 (first entry)
XX
DE PASK46-p14XH encoded C-terminal streptavidin-binding sequence.
XX
XX Streptavidin binding peptide; fusion protein; immunoglobulin; Ig;
XX heavy chain variable region; affinity chromatography; purification;
XX peptide tag.
XX
XX Synthetic.
OS
FH Key Location/Qualifiers
FT Peptide 4..11
FT /note= "streptavidin-binding octapeptide fused to C-terminus of VH chain"
FT
FT
PN GB2272698-A.
XX
PD 25-MAY-1994.
XX
PF 01-NOV-1993; 93GB-00022501.
XX
PR 03-NOV-1992; 92DE-04237113.
XX
PA (BIOA-) INST BIOANALYTIC GEMEINNETZIGE GMBH.
XX
XX Skerra A, Schmidt T;
PI WPI, 1994-153484/19.
XX
DR N-PSDB; AAQ62669.
XX
PT New fusion peptide(s) - have easily controlled binding properties and are
FT capable of binding to streptavidin.
XX
PS Disclosure; Page 11; 53pp; English.
XX
CC Six derivatives of PASK46 (a plasmid for expression of the D1.3FV
CC fragment in E.coli) were produced which encode 4 different peptides at
CC the C-terminus of VH and 2 peptides at the C-terminus of VL of the D1.3FV
CC fragment. The peptides fused to the C-terminus are all examples of
CC streptavidin-binding peptides corresponding to a generic formula (see
CC AAR52698). The peptides do not interfere with the protein function but
CC facilitate purification by conferring streptavidin-binding properties on

CC the fusion protein
XX
SQ Sequence 11 AA;

Query Match 100.0%; Score 48; DB 2; Length 11;
Best Local Similarity 87.5%; Pred. No. 0.13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 WKHPQFER 8
Db 4 WKHPQFER 11

RESULT 2
AAR52693
ID AAR52693 standard; protein; 11 AA.

AC AAR52693;
XX
DT 10-JAN-1995 (first entry)
XX
DE pASK46-pxlX encoded C-terminal streptavidin-binding sequence.
XX
KW Streptavidin binding peptide; fusion protein; immunoglobulin; Ig;
KW heavy chain variable region; affinity chromatography; purification;
KW peptide tag.
XX
OS Synthetic.

XX
FH Key Location/Qualifiers
FT Peptide 4..11
FT /note= "streptavidin-binding octapeptide fused to C-terminus of VH chain"

GB2272698-A.
XX
PD 25-MAY-1994.

XX
PF 01-NOV-1993; 93GB-00022501.
XX
PR 03-NOV-1992; 92DE-04237113.

XX
PA (BIOA-) INST BIOANALYTIK GEMEINNUTZIGE GMBH.

XX
PI Skerra A, Schmidt T;

XX
DR WPI; 1994-153484/19.
XX
DR N-PSDB; AAQ62668.

XX
PT New fusion peptide(s) - have easily controlled binding properties and are capable of binding to streptavidin.

XX
PS Disclosure; Page 11; 53pp; English.

XX
CC Six derivatives of pASK46 (a plasmid for expression of the D1.3Fv fragment in E.coli) were produced which encode 4 different peptides at the C-terminus of VH and 2 peptides at the C-terminus of VL of the D1.3Fv fragment. The peptides fused to the C-terminus are all examples of streptavidin-binding peptides corresponding to a generic formula (see AAR52698). The peptides do not interfere with the protein function but facilitate purification by conferring streptavidin-binding properties on the fusion protein

XX
SQ Sequence 11 AA;

Query Match 100.0%; Score 48; DB 2; Length 11;
Best Local Similarity 87.5%; Pred. No. 0.13;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 WKHPQFER 8
Db 4 WKHPQFER 11

RESULT 3
AAR52697
ID AAR52697 standard; protein; 12 AA.

XX
AC AAR52697;
XX
DT 10-JAN-1995 (first entry)
XX
DE pASK46-pxlX encoded C-terminal streptavidin-binding sequence.

XX
KW Streptavidin binding peptide; fusion protein; immunoglobulin; Ig;
KW light chain variable region; affinity chromatography; purification;
KW peptide tag.

XX
OS Synthetic.

XX
FH Key Location/Qualifiers
FT Peptide 5..12
FT /note= "streptavidin-binding octapeptide fused to C-terminus of VL chain"

GB2272698-A.

XX
PD 25-MAY-1994.

XX
PF 01-NOV-1993; 93GB-00022501.

XX
PR 03-NOV-1992; 92DE-04237113.

XX
PA (BIOA-) INST BIOANALYTIK GEMEINNUTZIGE GMBH.

XX
PI Skerra A, Schmidt T;

XX
DR WPI; 1994-153484/19.
XX
DR N-PSDB; AAQ62672.

XX
PT New fusion peptide(s) - have easily controlled binding properties and are capable of binding to streptavidin.

XX
PS Disclosure; Page 11; 53pp; English.

XX
CC Six derivatives of pASK46 (a plasmid for expression of the D1.3Fv fragment in E.coli) were produced which encode 4 different peptides at the C-terminus of VH and 2 peptides at the C-terminus of VL of the D1.3Fv fragment. The peptides fused to the C-terminus are all examples of streptavidin-binding peptides corresponding to a generic formula (see AAR52698). The peptides do not interfere with the protein function but facilitate purification by conferring streptavidin-binding properties on the fusion protein

XX
SQ Sequence 12 AA;

Query Match 100.0%; Score 48; DB 2; Length 12;
Best Local Similarity 87.5%; Pred. No. 0.15;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 WKHPQFER 8
Db 5 WKHPQFER 12

RESULT 4
AAM59212
ID AAM59212 standard; peptide; 8 AA.

XX
AC AAM59212;

XX
DT 27-AUG-1998 (first entry)

XX
DE Streptavidin tagged peptide ligand #2.

XX
KW Streptavidin; ligand; binding affinity; mutant; isolation; purification;

```

KW recover; immobilise.
XX
XX Synthetic.
XX
XX EP835934-A2.
XX
XX 15-APR-1998.
XX
XX
XX 09-OCT-1997; 97EP-00117504.
XX
XX 10-OCT-1996; 96DE-01041876.
XX
XX (BIOA-) INST BIOANALYTIK GMBH.
XX
XX Skerra A, Voess S;
XX
XX WPI; 1998-218668/20.
XX
XX Streptavidin mutants with higher binding affinity for peptide ligands -
XX PT have mutation in amino acid region 44-53, used to isolate, purify or
XX PT determine fusion proteins including these ligands.
XX
XX Claim 10; Page 11; 21pp; German.
XX
XX AAW59211 and AAW59212 are ligands used in a method to assay binding
XX CC affinity of streptavidin mutants. These mutants have a mutation within
XX CC the amino acid (aa) region 44-53 of the wild-type protein show a higher
XX CC binding affinity than the wild-type for peptide ligands that include the
XX CC sequence of formula  $\text{Trp-X-His-Pro-Gln-Phe-Y-Z}$  where X = any aa; Y and Z
XX CC are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin
XX CC mutants can be used to isolate, purify and determine proteins or to
XX CC determine/recover substances that contain streptavidin-binding groups.
XX CC Such compounds may also be used to immobilise fusions on microtitre
XX CC plates, microbeads or sensor chips
XX
XX
SQ Sequence 8 AA;

Query Match 93.8%; Score 45; DB 2; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.8e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFER 8
   |:|||||:
   1 WSHPOFEK 8

Db WO200114578-A1.

RESULT 5
AAB35433
ID AAB35433 standard; peptide; 8 AA.
XX
XX AAB35433;
XX
XX 23-MAY-2001 (first entry)
XX
XX Epitope peptide #3.
XX
XX Nascent protein detection; protein analysis; aminoacylated tRNA;
XX KM BODIPY marker; disease diagnosis.
XX
XX Unidentified.
XX
XX WO200114578-A1.
XX
XX 01-MAR-2001.
XX
XX 23-AUG-2000; 2000WO-US023233.
XX
XX 25-AUG-1999; 99US-00382736.
XX
XX 25-AUG-1999; 99US-00382950.
XX
XX (AMBE-) AMBERGEN INC.
XX
XX Rothenchild KJ, Gite S, Olejnik J;
XX

```

```

XX
XX WPI; 2001-168972/17.
XX
XX Method for detecting nascent proteins by fluorescence comprises
XX PT misaminoacylating a tRNA molecule with a marker compound, useful for
XX PT detecting mutations in proteins, e.g. cancer.
XX
XX
XX Disclosure; Page 47; 204pp; English.
XX
XX The present invention describes a method of detecting nascent proteins
XX CC involving aminoacylating a tRNA molecule with a 4,4-difluoro-4-bora-3A,4A
XX CC -diaz-a-8-indacene (BODIPY) marker leading to the production of a
XX CC misaminoacylated tRNA. This enables the detection, isolation and analysis
XX CC of nascent proteins using UV without the usual accompanying radioactivity
XX CC problems. It may be used to detect mutations, for example in cancer.
XX CC Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer
XX
XX
SQ Sequence 8 AA;

Query Match 93.8%; Score 45; DB 4; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.8e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WKHPQFER 8
   |:|||||:
   1 WSHPOFEK 8

Db WO200114578-A1.

RESULT 6
AAB35442
ID AAB35442 standard; peptide; 8 AA.
XX
XX AAB35442;
XX
XX 23-MAY-2001 (first entry)
XX
XX Nascent protein detection method related peptide #4.
XX
XX Nascent protein detection; protein analysis; aminoacylated tRNA;
XX KM BODIPY marker; disease diagnosis.
XX
XX Unidentified.
XX
XX WO200114578-A1.
XX
XX 01-MAR-2001.
XX
XX 23-AUG-2000; 2000WO-US023233.
XX
XX 25-AUG-1999; 99US-00382736.
XX
XX 25-AUG-1999; 99US-00382950.
XX
XX (AMBE-) AMBERGEN INC.
XX
XX Rothenchild KJ, Gite S, Olejnik J;
XX
XX WPI; 2001-168972/17.
XX
XX Method for detecting nascent proteins by fluorescence comprises
XX PT misaminoacylating a tRNA molecule with a marker compound, useful for
XX PT detecting mutations in proteins, e.g. cancer.
XX
XX
XX Example 22; Page 153; 204pp; English.
XX
XX The present invention describes a method of detecting nascent proteins
XX CC involving aminoacylating a tRNA molecule with a 4,4-difluoro-4-bora-3A,4A
XX CC -diaz-a-8-indacene (BODIPY) marker leading to the production of a
XX CC misaminoacylated tRNA. This enables the detection, isolation and analysis
XX CC of nascent proteins using UV without the usual accompanying radioactivity
XX CC problems. It may be used to detect mutations, for example in cancer.
XX CC Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer
XX
XX
SQ Sequence 8 AA;

```

Query Match 93.8%; Score 45; DB 4; Length 8;
 Best Local Similarity 75.0%; Pred. No. 1.8e+06;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFER 8
 DB 1 WSHPOFEK 8

RESULT 7

AB68616
 ID AAB68616 standard; peptide; 8 AA.

AC AAB68616;

DT 27-APR-2001 (first entry)

DE Strep-Tag II sequence.

KM Protein-RNA fusion; Strep-Tag II.

OS Unidentified.

PN WO200107657-A1.

PD 01-FEB-2001.

PF 19-JUL-2000; 2000WO-US019653.

PR 27-JUL-1999; 99US-0145834P.

PA (PHYL-) PHYLLOS INC.

PI Kurz M, Lohse P, Wagner R;

DR WPI; 2001-182803/18.

PT Affixing a peptide acceptor to an RNA molecule useful for producing
 PT fusion proteins for isolating proteins or nucleic acids with desired
 PT properties through attachment of a peptide acceptor to the 3' end of an
 PT RNA molecule.

PS Example 5; Page 22; 56pp; English.

CC The present invention relates to a method for affixing a peptide acceptor
 CC to an RNA molecule through the formation of a covalent bond, noncovalent
 CC bond, or by chemical ligation. The method is useful for producing RNA-
 CC protein fusions which can be used for the isolation of proteins or
 CC nucleic acids with desired properties from large pools of partially or
 CC completely random amino acid or nucleic acid sequences. The present
 CC sequence is a Strep-Tag II used in the present invention

SO Sequence 8 AA;

Query Match 93.8%; Score 45; DB 4; Length 8;
 Best Local Similarity 75.0%; Pred. No. 1.8e+06;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFER 8
 DB 1 WSHPOFEK 8

RESULT 8

AA019065
 ID AA019065 standard; peptide; 8 AA.

AC AA019065;

DT 14-NOV-2002 (first entry)

DE Mutation detection method tag peptide SEQ ID NO: 34.

XX Mutation detection; primer; mutant; tag; tumour suppressor gene;
 KM protein production; cancer.

OS Synthetic.

PN WO200266675-A2.

PD 29-AUG-2002.

PF 15-FEB-2002; 2002WO-EP001651.

PR 16-FEB-2001; 2001DE-01007317.

PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PI Kahmann S, Mueller O;

DR WPI; 2002-674959/72.

DR N-PSDB; AAL49464.

PT Detecting mutations in nucleic acid, useful for diagnosis and
 PT characterization of tumors, by amplification, in vitro transcription and
 PT translation, then protein detection.

PS Disclosure; Fig 5; 62pp; German.

CC The present invention relates to a method of detecting mutations in a
 CC nucleic acid by amplifying the nucleic acid to produce a double-stranded
 CC amplicon, in vitro transcription and translation of this amplicon, and
 CC detection of the translated protein. The primers used for amplification
 CC are designed to produce an amplicon that is translatable and allows
 CC differentiation between translation products of wild-type and mutated
 CC nucleic acids. The method is used to detect mutations in tumour
 CC suppressor genes, for (early) diagnosis, monitoring and characterisation
 CC of tumours (especially of bladder and intestine) and in the germ line
 CC (using nucleic acids from embryos or blood cells). A new multi-tag vector
 CC is used to detect or verify the reading frame of a nucleic acid cloned in
 CC it, and to determine the suitability of detectable peptides for analysis
 CC and/or purification of a recombinant protein, expressed from a sequence
 CC cloned in the vector. The present sequence is a tag peptide which was
 CC used in the invention

SO Sequence 8 AA;

Query Match 93.8%; Score 45; DB 5; Length 8;
 Best Local Similarity 75.0%; Pred. No. 1.8e+06;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXHPOFER 8
 DB 1 WSHPOFEK 8

RESULT 9

ABB57464
 ID ABB57464 standard; peptide; 8 AA.

AC ABB57464;

DT 18-MAR-2002 (first entry)

DE STREP tag II peptide.

KM Immunomodulatory human MHC class II antigen-binding protein, HLA;
 KW human leukocyte antigen; immune system; immunosuppression; antibody;
 KW major histocompatibility complex; antirheumatic; antiarthritic;
 KW neuroprotective; antiinflammatory; antidiabetic; antiparrotic;
 KW immunosuppressive; dermatological; antithyroid; nephrotoxic; psoriasis;
 KW thymomimetic; hepatotropic; immune response suppressor; narcolepsy;
 KW rheumatoid arthritis; juvenile arthritis; multiple sclerosis; insulinitis;
 KW Grave's disease; insulin-dependent diabetes; Hashimoto's disease;
 KW systemic lupus erythematosus; ankylosing spondylitis; myasthenia gravis;

KW transplant rejection; graft versus host disease; pemphigus vulgaris;
 KW glomerulonephritis; thyroiditis; pancreatitis; primary biliary cirrhosis;
 KW irritable bowel disease; Sjogren's syndrome.
 XX
 OS Synthetic.
 XX
 PN WO200187338-A1.
 XX
 PD 22-NOV-2001.
 XX
 PF 14-MAY-2001; 2001WO-US015626.
 XX
 PR 12-MAY-2000; 2000EP-00110063.
 PR 06-OCT-2000; 2000US-0238762P.
 XX
 PA (GPCB-) GPC BIOTECH AG.
 PA (MORP-) MORPHOSYS AG.
 XX
 PI Nagy Z, Tesar M, Thomassen-Wolf B;
 XX
 DR WPI; 2002-075289/10.
 XX
 PT Composition for suppressing immune response, treating diseases of immune
 PT system, has polypeptide comprising antibody-based antigen-binding domain
 PT of human composition, which binds antigen expressed on a cell surface.
 XX
 PS Disclosure; Page 31; 139pp; English.
 XX
 CC The present invention describes a composition (I), comprising a
 CC polypeptide comprising an antibody-based antigen-binding domain of human
 CC composition with binding specificity for an antigen expressed on the
 CC surface of a cell, where treating cells expressing the antigen with the
 CC polypeptides leads to suppression of an immune response, and the IC50 for
 CC the suppression of immune response is 1 microm or less. (I) has
 CC antirheumatic, antiarthritic, neuroprotective, antiinflammatory,
 CC antidiabetic, nephroprotective, immunosuppressive, dermatological,
 CC anticholesteric, neurotropic, chymolytic and hepatotropic activities, and
 CC can be used as a suppressor of immune response (I) is useful for
 CC suppressing activation or proliferation of a cell of the immune system,
 CC suppressing IL-2 secretion by a cell, the interaction of a cell of the
 CC immune system with another cell, immunosuppressing a patient and for
 CC killing a cell expressing an antigen, human leukocyte antigen (HLA)-DR on
 CC the surface of the cell, where neither cytotoxic entities nor
 CC immunological mechanisms are needed to cause or lead to the killing. (I)
 CC (optionally linked to cytotoxic or immunogenic agent) is useful for
 CC preparing a pharmaceutical preparation for the treatment of rheumatoid
 CC arthritis, juvenile arthritis, multiple sclerosis, Grave's disease,
 CC insulin-dependent diabetes, narcolepsy, psoriasis, systemic lupus
 CC erythematosus, ankylosing spondylitis, transplant rejection, graft versus
 CC host disease, Hashimoto's disease, myasthenia gravis, pemphigus vulgaris,
 CC glomerulonephritis, thyroiditis, pancreatitis, insulinitis, primary biliary
 CC cirrhosis, irritable bowel disease and Sjogren's syndrome in humans.
 CC ABA92469 to ABA92474 and ABB57457 to ABB57590 represent sequence used in
 CC the exemplification of the present invention
 XX
 SQ Sequence 8 AA:
 Query Match 93.8%; Score 45; DB 5; Length 8;
 Best Local Similarity 75.0%; Pred. No. 1.8e+06;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WSHPOFER 8
 DB 1 WSHPOFER 8
 RESULT 10
 ID ABB77486 standard; peptide; 8 AA.
 XX ABB77486;
 AC
 XX 22-JUL-2002 (first entry)
 DT

XX
 DE AAV-helper plasmid related Strep-peptide tag SEQ ID NO 17.
 XX
 KW Adeno-associated virus; AAV; glycoprotein; ITR; cap; rep; vector.
 XX
 OS Synthetic.
 XX
 PN WO200238782-A2.
 XX
 PD 16-MAY-2002.
 XX
 PF 13-NOV-2001; 2001WO-EP013125.
 XX
 PR 13-NOV-2000; 2000DE-01056210.
 XX
 PA (ARIM-) ARIMEDES BIOTECHNOLOGY GMBH.
 PA Orberger G, Hellmuth K, Wagener C;
 XX
 PI WPI; 2002-435853/46.
 XX
 DR
 XX
 PT Vector system for preparing recombinant adeno-associated viral particles,
 PT used for high-level expression of heterologous therapeutic proteins in
 PT eukaryotic cells.
 XX
 PS Disclosure; Page 17; 59pp; German.
 XX
 CC The invention relates to a viral vector system for preparing recombinant
 CC adeno-associated virus (AAV) particles comprising: at least two plasmid
 CC vectors (ABL58983 and ABL58984) that include the two inverted terminal
 CC repeats (ITR) of AAV and additional sequences and plasmid vectors without
 CC ITRs but containing the rep and cap genes of AAV required for replication
 CC and packaging. The system is useful for producing recombinant AAV for
 CC production of a wide range of therapeutic glycoproteins in eukaryotic
 CC cells. The system provides efficient, large scale production of
 CC heterologous proteins in mammalian cells, without requiring an adenovirus
 CC helper. It is not toxic to host cells and does not cause lysis, so
 CC produced proteins are highly pure. The present sequence is that of a
 CC peptide tag, useful to the invention
 XX
 SQ Sequence 8 AA:
 Query Match 93.8%; Score 45; DB 5; Length 8;
 Best Local Similarity 75.0%; Pred. No. 1.8e+06;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 WSHPOFER 8
 DB 1 WSHPOFER 8
 RESULT 11
 ID ABB73584 standard; protein; 8 AA.
 XX ABB73584;
 AC
 XX 03-MAR-2003 (first entry)
 DT
 XX
 DE StreptagII affinity peptide.
 XX
 KW Non-protein L-amino acid; O-acetyl-L-serine; agrochemical production;
 KW nucleophile; O-acetyl-L-serine sulphhydriylase; pharmaceutical production.
 XX
 OS Synthetic.
 XX
 PN EP1247869-A1.
 XX
 PD 09-OCT-2002.
 XX
 PF 28-MAR-2002; 2002EP-00007262.
 XX
 AC
 XX 04-APR-2001; 2001DE-01016881.
 DT

PR 03-MAY-2001; 2001DE-01021515.
XX
PA (CONE) CONSORTIUM ELEKTROCHEM IND GMBH.
XX
PI Maier T, Gaebert C;
XX
DR WPI; 2003-077522/08.
XX
PT Production of non-protein L-amino acids useful for the manufacture of
PT pharmaceuticals and agrochemicals, comprises an enzyme catalyzed reaction
PT of O-acetyl-L-serine with a nucleophile.
XX
PS Example 1; Page 8; 20pp; German.
XX
CC This invention describes a novel method for the production of non-protein
CC L-amino acids by the reaction of O-acetyl-L-serine with a nucleophile in
CC the presence of a catalyst comprising O-acetyl-L-serine sulphydrylase at
CC pH 5-7.4. The method of the invention is useful for the manufacture of
CC pharmaceuticals and agrochemicals. In contrast to the process described
CC in DE 10046934, a high nucleophile concentration can be used which
CC includes toxic compounds. This sequence represents an affinity peptide
CC containing a StreptagII motif which is used in the construction of fusion
CC genes containing *Escherichia coli* cysK and cysM fragments
XX
SQ Sequence 8 AA;
XX
Query Match 93.8%; Score 45; DB 6; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.8e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 WXHPQFER 8
DB 1 WSHPOFEK 8
XX
RESULT 12
AAE37230
ID AAE37230 standard; peptide; 8 AA.
XX
AC AAE37230;
XX
DT 07-AUG-2003 (first entry)
XX
DE Strept-tag II epitope peptide.
XX
KW Gene expression; therapy; isolation; epitope.
XX
OS Synthetic.
XX
PN WO2003038049-A2.
XX
PD 08-MAY-2003.
XX
PF 29-OCT-2002; 2002MO-US034645.
XX
PR 29-OCT-2001; 2001US-0340689P.
XX
PA (RENO-) RENOVIS INC.
XX
PI Heintz N, Serafini TA, Shyjan AW;
XX
DR WPI; 2003-430512/40.
XX
PT Isolating cell-type specific mRNAs, useful in gene expression analysis or
PT quantification in a specific cell in a heterogeneous cell mixture, by
PT isolating mRNA bound to ribosomes or proteins that bind mRNA in a cell
PT type specific manner.
XX
PS Example 2; Page 121; 136pp; English.
XX
CC The invention relates to a method for isolating mRNA from a population of
CC cells. The method involves selectively isolating ribosomes or proteins
CC that bind mRNA in a cell type specific manner and then isolating the mRNA

CC bound to the ribosomes or proteins that bind mRNA. The method is useful
CC for facilitating the analysis and quantification of gene expression in a
CC selected cell type present within a heterogeneous cell mixture. The
CC method may also be used in diagnostics or therapies for human diseases.
CC The present sequence is Strept-tag II epitope peptide. This sequence is
CC used to illustrate the method of the invention
XX
SQ Sequence 8 AA;
XX
Query Match 93.8%; Score 45; DB 6; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.8e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 WXHPQFER 8
DB 1 WSHPOFEK 8
XX
RESULT 13
ABP60361
ID ABP60361 standard; peptide; 8 AA.
XX
AC ABP60361;
XX
DT 28-MAR-2003 (first entry)
XX
DE Streptavidin 11 tag peptide SEQ ID NO 2.
XX
KW Streptavidin; protein chip; microtitre plate; detection.
XX
OS Synthetic.
XX
PN DE10113776-A1.
XX
PD 02-OCT-2002.
XX
PF 21-MAR-2001; 2001DE-01013776.
XX
PR 21-MAR-2001; 2001DE-01013776.
XX
PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
PI Schmidt T;
XX
DR WPI; 2003-031166/03.
XX
PT New isolated peptide, useful as affinity purification tag for recombinant
PT protein, comprises at least two high-affinity streptavidin-binding
PT modules.
XX
PS Example 1; Page 8; 18pp; German.
XX
CC The invention relates to an isolated peptide (I) comprising at least two
CC individual modules separated by 0-50 amino acids, with each containing at
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
CC streptavidin binding modules, are useful as affinity handles for
CC purification of recombinant fusion proteins (FP), also for detecting FP,
CC e.g. on protein chips or microtitre plates. The modules in (I) bind
CC strongly to streptavidin, with a co-operative effect that provides
CC stronger binding than a single tag, but are displaced by a competitor.
CC (I) does not interfere with the function of attached proteins (II) (so it
CC may not be essential to remove it); facilitates detection and has easily
CC controllable binding properties. (I) is particularly used for purifying
CC FP from dilute solution in batch formats (which use simpler apparatus
CC than column methods and result in lower loss of FP). The present sequence
CC is that of a streptavidin tag peptide disclosed with the invention
XX
SQ Sequence 8 AA;
XX
Query Match 93.8%; Score 45; DB 6; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.8e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

[illegible]

DE Steptag epitope useful as an affinity marker.

KX Non-radioactive marker; nascent protein detection method;
KW cellular translation system; cell-free translation system;
KM dipyrrometheneboron difluoride dye;
KV 4,4-difluoro-4-bora-3a,4a-diaza-s-indacene dye; vaccine; drug;
KW human disease screening; human disorder; protein separation;
KW affinity marker; Steptag epitope.

OS Synthetic.
XX
PN US2003092031-A1.
XX
PD 15-MAY-2003.
XX
PF 18-JUN-2002; 2002US-00174368.
XX
PR 25-AUG-1999; 99US-00382736.
PR 23-AUG-2000; 2000WO-US023233.
PR 21-JUN-2002; 2002US-00049332.
PA (AMBE-) AMBERGEN INC.
PI Rothschild KJ, Gite S, Olejnik J;
PI WPI; 2003-576764/54.
DR XX
PT Detecting, analyzing or isolating nascent proteins comprises introducing
PT a modified nucleic acid template into a cellular or cell-free translation
PT system to generate a nascent protein having at least an N-terminal
PT marker.
XX
PS Disclosure; Page 15; 76pp; English.

CC The present invention relates to non-radioactive markers used in the
CC detection and analysis of nascent proteins translated in cellular or cell
CC free translation systems. The preferred non-radioactive markers are
CC dipyrrometheneboron difluoride (4,4-difluoro-4-bora-3a,4a-diaza-s-
CC indacene) dyes. The detection method of the invention is a gel-free
CC method that comprises introducing a modified nucleic acid template into a
CC translation system under conditions such that a nascent protein is
CC generated, the protein comprising at least an N-terminal marker. The
CC method is useful in detecting, analyzing and isolating nascent proteins
CC produced in a cell-free or cellular translation system without the use of
CC radioactive amino acids or other radioactive labels. Compositions
CC comprising nascent proteins translated in the presence of markers may be
CC used as vaccines or as drugs for humans and other animals. The method and
CC a kit containing reagents for the detection of nascent proteins may be
CC used as a rapid means to screen humans or other animals for the presence
CC of certain diseases or disorders. The present sequence represents an
CC epitope that can be used as an affinity marker for protein separation.
XX

SQ Sequence 8 AA;

Query Match 93.8%; Score 45; DB 7; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.8e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0.

QY 1 WXHPDPER 8
|:||||:
Db 1 WSHDPDFK 8

Search completed: March 2, 2005, 13:02:44
Job time : 36.0325 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 24.878 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ3

Perfect score: 48

Sequence: 1 wxhpqfer 8

Scoring table:

BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 32804528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	93.8	8	9	US-09-809-517A-9
2	45	93.8	8	9	US-09-973-145-7
3	45	93.8	8	12	US-09-813-197-8
4	45	93.8	8	13	US-10-104-218-5
5	45	93.8	8	13	US-10-208-357-9
6	45	93.8	8	14	US-10-001-934-8
7	45	93.8	8	14	US-10-026-578B-2
8	45	93.8	8	14	US-10-026-578B-9
9	45	93.8	8	14	US-10-174-368A-7
10	45	93.8	8	14	US-10-264-127-8
11	45	93.8	8	15	US-10-339-712-8
12	45	93.8	8	15	US-10-339-712-67
13	45	93.8	8	15	US-10-275-046-4

14	45	93.8	8	15	US-10-425-000-76	Sequence 76, Appl
15	45	93.8	8	15	US-10-424-999-25	Sequence 25, Appl
16	45	93.8	8	15	US-10-358-283-23	Sequence 23, Appl
17	45	93.8	8	16	US-10-628-432-41	Sequence 41, Appl
18	45	93.8	8	17	US-10-494-248-17	Sequence 17, Appl
19	45	93.8	8	17	US-10-634-645-11	Sequence 11, Appl
20	45	93.8	8	17	US-10-719-523-8	Sequence 8, Appl
21	45	93.8	9	9	US-09-983-067-3	Sequence 3, Appl
22	45	93.8	10	9	US-09-809-517A-6	Sequence 6, Appl
23	45	93.8	10	15	US-10-147-211A-20	Sequence 20, Appl
24	45	93.8	11	15	US-10-354-983-29	Sequence 29, Appl
25	45	93.8	11	16	US-10-628-432-25	Sequence 25, Appl
26	45	93.8	21	9	US-09-809-517A-30	Sequence 30, Appl
27	45	93.8	22	9	US-09-809-517A-33	Sequence 33, Appl
28	45	93.8	24	9	US-09-809-517A-31	Sequence 31, Appl
29	45	93.8	24	14	US-10-026-578B-3	Sequence 3, Appl
30	45	93.8	24	14	US-10-026-578B-4	Sequence 4, Appl
31	45	93.8	25	9	US-09-809-517A-34	Sequence 34, Appl
32	45	93.8	36	14	US-10-026-578B-10	Sequence 10, Appl
33	45	93.8	36	14	US-10-026-578B-11	Sequence 11, Appl
34	45	93.8	117	10	US-09-977-137A-4	Sequence 4, Appl
35	45	93.8	117	10	US-09-977-137A-5	Sequence 5, Appl
36	45	93.8	117	10	US-09-977-137A-7	Sequence 7, Appl
37	45	93.8	117	10	US-09-977-137A-8	Sequence 8, Appl
38	45	93.8	117	10	US-09-977-137A-9	Sequence 9, Appl
39	45	93.8	117	10	US-09-977-137A-10	Sequence 10, Appl
40	45	93.8	117	10	US-09-977-137A-11	Sequence 11, Appl
41	45	93.8	117	10	US-09-977-137A-12	Sequence 12, Appl
42	45	93.8	118	10	US-09-977-137A-6	Sequence 6, Appl
43	45	93.8	245	17	US-10-887-228A-1	Sequence 1, Appl
44	45	93.8	246	17	US-10-887-228A-9	Sequence 9, Appl
45	45	93.8	252	17	US-10-887-228A-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-809-517A-9
Sequence 9, Application US/09809517A
Patent No. US20020034733A1
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: Particles via disulfide bonds
FILE REFERENCE: MORPHO/1
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 9
LENGTH: 8
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-9

Query Match
Best Local Similarity 93.8%; Score 45; DB 9; Length 8;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 wxhpqfer 8
|:|||||:
Db 1 wxhpqfer 8

RESULT 2
US-09-973-145-7

```
Sequence 7, Application US/09973145
Patent No. US20020132248A1
GENERAL INFORMATION:
APPLICANT: Rothschild, Kenneth J.
APPLICANT: Gile, Sadanand
APPLICANT: Olejnik, Jerzy
FILE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
FILE REFERENCE: AMBER-06819
CURRENT APPLICATION NUMBER: US/09/973,145
CURRENT FILING DATE: 2001-10-09
PRIOR APPLICATION NUMBER: 09/382,950
PRIOR FILING DATE: 1999-08-25
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Synthetic
US-09-973-145-7
```

```
Query Match          93.8%; Score 45; DB 9; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXPQFER 8
        |||||||:
Db       1 WSHPOFEK 8
```

```
RESULT 3
US-09-813-197-8
Sequence 8, Application US/09813197
Publication No. US2005009013A1
GENERAL INFORMATION:
APPLICANT: Rothschild, Kenneth
APPLICANT: Gile, Sadanand
APPLICANT: Olejnik, Jerzy
FILE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Pro
FILE REFERENCE: AMBER-03951
CURRENT APPLICATION NUMBER: US/09/813,197
CURRENT FILING DATE: 2001-03-20
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.0
SEQ ID NO 8
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-09-813-197-8
```

```
Query Match          93.8%; Score 45; DB 12; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXPQFER 8
        |||||||:
Db       1 WSHPOFEK 8
```

```
RESULT 4
US-10-104-218-5
Sequence 5, Application US/10104218
Publication No. US20020177196A1
GENERAL INFORMATION:
APPLICANT: MAIER, Thomas
APPLICANT: GABBERT, Carsten
FILE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
```

```
FILE REFERENCE: MAIER, T. ET AL.-2
CURRENT APPLICATION NUMBER: US/10/104,218
CURRENT FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 16 881.0
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 21 515.0
PRIOR FILING DATE: 2001-05-03
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)..(8)
OTHER INFORMATION: StreptagII affinity peptide for protein purification
US-10-104-218-5
```

```
Query Match          93.8%; Score 45; DB 13; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXPQFER 8
        |||||||:
Db       1 WSHPOFEK 8
```

```
RESULT 5
US-10-208-357-9
Sequence 9, Application US/10208357
Publication No. US20020182687A1
GENERAL INFORMATION:
APPLICANT: Kutz, Markus
APPLICANT: Lohse, Peter
APPLICANT: Magnet, Richard
FILE OF INVENTION: Peptide Acceptor Ligation Methods
FILE REFERENCE: 50036/031002
CURRENT APPLICATION NUMBER: US/10/208,357
CURRENT FILING DATE: 2002-07-30
PRIOR APPLICATION NUMBER: US/09/619,103
PRIOR FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 60/145,834
PRIOR FILING DATE: 1999-07-27
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: designed sequence to act as an identifying tag
US-10-208-357-9
```

```
Query Match          93.8%; Score 45; DB 13; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 WXPQFER 8
        |||||||:
Db       1 WSHPOFEK 8
```

```
RESULT 6
US-10-001-934-8
Sequence 8, Application US/10001934
Publication No. US20030032782A1
GENERAL INFORMATION:
APPLICANT: NAGY, ZOLTAN
FILE OF INVENTION: HUMAN PEPTIDES/PROTEINS CAUSING OR LEADING TO THE
FILE REFERENCE: GRCG-P01-003
CURRENT APPLICATION NUMBER: US/10/001,934
```

```

; CURRENT FILING DATE: 2001-11-15
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: STREP tag
US-10-001-934-8
```

```

Query Match          93.8%; Score 45; DB 14; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 WKHPQFER 8
        |:|||||:
Db      1 WSHPOFEK 8
```

```

RESULT 7
US-10-026-578B-2
; Sequence 2, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
; NAME/KEY: misc.feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-2
```

```

Query Match          93.8%; Score 45; DB 14; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 WKHPQFER 8
        |:|||||:
Db      1 WSHPOFEK 8
```

```

RESULT 8
US-10-026-578B-9
; Sequence 9, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
```

```

; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic peptide
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-9
```

```

Query Match          93.8%; Score 45; DB 14; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 WKHPQFER 8
        |:|||||:
Db      1 WSHPOFEK 8
```

```

RESULT 9
US-10-174-368A-7
; Sequence 7, Application US/10174368A
; Publication No. US20030092031A1
; GENERAL INFORMATION:
; APPLICANT: Rochechild, Kenneth J.
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prc
; FILE REFERENCE: AMBR-07145
; CURRENT APPLICATION NUMBER: US/10/174,368A
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: 10/049,332
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/23233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; NAME/KEY: misc.feature
; OTHER INFORMATION: Synthetic
US-10-174-368A-7
```

```

Query Match          93.8%; Score 45; DB 14; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 WKHPQFER 8
        |:|||||:
Db      1 WSHPOFEK 8
```

```

RESULT 10
US-10-264-127-8
; Sequence 8, Application US/10264127
; Publication No. US20030190643A1
; GENERAL INFORMATION:
; APPLICANT: Rochechild, Kenneth
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot
; FILE REFERENCE: AMBR-03951
; CURRENT APPLICATION NUMBER: US/10/264,127
```

```
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: US/09/382,736B
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-264-127-8
```

```
Query Match      93.8%; Score 45; DB 14; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 WXHPOFER 8
Db      1 WSHPOFEK 8
```

```
RESULT 11
US-10-339-712-8
; Sequence 8, Application US/10339712
; Publication No. US20040014071A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Methode for the Detection, Analysis and Isolation of Nascent
; FILE REFERENCE: Proteins
; CURRENT APPLICATION NUMBER: US/10/339,712
; PRIOR FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: 10/049,322
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/23233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-339-712-8
```

```
Query Match      93.8%; Score 45; DB 15; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 WXHPOFER 8
Db      1 WSHPOFEK 8
```

```
RESULT 12
US-10-339-712-67
; Sequence 67, Application US/10339712
; Publication No. US20040014071A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent
; FILE REFERENCE: Proteins
; OTHER INFORMATION: Synthetic
US-10-339-712-67
```

```
; CURRENT APPLICATION NUMBER: US/10/339,712
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: 10/049,322
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/23233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 67
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-339-712-67
```

```
Query Match      93.8%; Score 45; DB 15; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 WXHPOFER 8
Db      1 WSHPOFEK 8
```

```
RESULT 13
US-10-275-046-4
; Sequence 4, Application US/10275046
; Publication No. US20040019187A1
; GENERAL INFORMATION:
; APPLICANT: Nagy et al.
; TITLE OF INVENTION: IMMUNOMODULATORY HUMAN MHC CLASS II ANTIGEN-BINDING POLYPEPTIDES
; FILE REFERENCE: GPGC-P01-260
; CURRENT APPLICATION NUMBER: US/10/275,046
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 00110063.5
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: STREP tag II
US-10-275-046-4
```

```
Query Match      93.8%; Score 45; DB 15; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 WXHPOFER 8
Db      1 WSHPOFEK 8
```

```
RESULT 14
US-10-425-000-76
; Sequence 76, Application US/10425000
; Publication No. US20040052777A1
; GENERAL INFORMATION:
; APPLICANT: Nesbitt, Mark
; APPLICANT: Cameron, Beatrice
; APPLICANT: Blanche, Francis
; TITLE OF INVENTION: Kringles Polypeptides and Methods for Using Them to Inhibit
; FILE REFERENCE: ST01027-B
; CURRENT APPLICATION NUMBER: US/10/425,000
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: 10/233,675
; PRIOR FILING DATE: 2002-09-04
```

/ NUMBER OF SEQ ID NOS: 105
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 76
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Purification tag
US-10-425-000-76

Query Match 93.8%; Score 45; DB 15; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXHPOFER 8
|:|||||:
Db 1 WSHPOFER 8

RESULT 15
US-10-424-999-25
/ Sequence 25, Application US/10424999
/ Publication No. US20040052810A1
/ GENERAL INFORMATION:
/ APPLICANT: Nesbitt, Mark
/ APPLICANT: Cameron, Beatrice
/ APPLICANT: Branche, Francis
/ TITLE OF INVENTION: Abrogen Polypeptides, Nucleic Acids Encoding Them and Methods for
/ FILE REFERENCE: ST01027-A
/ CURRENT APPLICATION NUMBER: US/10/424,999
/ PRIOR FILING DATE: 2003-04-29
/ PRIOR APPLICATION NUMBER: 10/233,675
/ NUMBER OF SEQ ID NOS: 70
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 25
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Purification tag
US-10-424-999-25

Query Match 93.8%; Score 45; DB 15; Length 8;
Best Local Similarity 75.0%; Pred. No. 1.2e+06;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WXHPOFER 8
|:|||||:
Db 1 WSHPOFER 8

Search completed: March 2, 2005, 14:18:49
Job time : 24.878 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 9.10569 Seconds
(without alignments)
65.585 Million cell updates/sec

Title: SEQ3

Perfect score: 48

Sequence: 1 wwhpqr 8

Scoring table:

BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	45	93.8	8	3	US-08-948-097-2
2	45	93.8	8	3	US-09-382-950-7
3	45	93.8	8	3	US-09-382-7368-8
4	45	93.8	8	4	US-09-619-103-9
5	45	93.8	8	4	US-10-104-218-5
6	45	93.8	8	4	US-09-809-517A-9
7	45	93.8	10	4	US-09-809-517A-6
8	45	93.8	10	4	US-09-809-517A-30
9	45	93.8	22	4	US-09-809-517A-33
10	45	93.8	24	4	US-09-809-517A-31
11	45	93.8	25	4	US-09-809-517A-34
12	45	93.8	117	4	US-09-977-137A-5
13	45	93.8	117	4	US-09-977-137A-7
14	45	93.8	117	4	US-09-977-137A-8
15	45	93.8	117	4	US-09-977-137A-9
16	45	93.8	117	4	US-09-977-137A-10
17	45	93.8	117	4	US-09-977-137A-11
18	45	93.8	117	4	US-09-977-137A-12
19	45	93.8	117	4	US-09-977-137A-13
20	45	93.8	118	4	US-09-977-137A-14
21	39	81.2	207	3	US-09-302-769-36
22	38	79.2	8	3	US-08-948-097-16
23	38	79.2	9	1	US-08-660-626-3
24	38	79.2	9	3	US-08-828-741B-14
25	38	79.2	9	3	US-08-948-097-1
26	38	79.2	9	3	US-09-031-168-3
27	38	79.2	9	3	US-09-160-567-14

28	38	79.2	9	4	US-09-672-239-3	Sequence 3, Appl
29	38	79.2	9	4	US-09-710-299-14	Sequence 14, Appl
30	38	79.2	9	4	US-09-509-031-14	Sequence 14, Appl
31	38	79.2	9	4	US-09-669-516C-3	Sequence 3, Appl
32	38	79.2	10	1	US-08-294-386C-11	Sequence 11, Appl
33	38	79.2	10	3	US-08-737-316A-5	Sequence 6, Appl
34	38	79.2	10	3	US-08-897-020-6	Sequence 11, Appl
35	38	79.2	10	3	US-08-895-707-11	Sequence 6, Appl
36	38	79.2	10	3	US-09-350-823-6	Sequence 13, Appl
37	38	79.2	10	4	US-09-715-805-13	Sequence 9, Appl
38	38	79.2	10	4	US-09-692-945-9	Sequence 7, Appl
39	38	79.2	10	4	US-09-640-041-7	Sequence 11, Appl
40	38	79.2	10	5	PCT-US95-10224-11	Sequence 17, Appl
41	38	79.2	12	1	US-08-737-757-17	Sequence 8, Appl
42	38	79.2	12	3	US-08-218-368-8	Sequence 8, Appl
43	38	79.2	12	4	US-09-904-599A-8	Sequence 8, Appl
44	38	79.2	12	5	PCT-US95-03742-8	Sequence 8, Appl
45	38	79.2	15	1	US-08-664-449-33	Sequence 33, Appl

ALIGNMENTS

```
RESULT 1
US-08-948-097-2
; Sequence 2, Application US/08948097C
; Patent No. 6103493
; GENERAL INFORMATION:
; APPLICANT: Skeira, Arne
; APPLICANT: Voss, Selma
; TITLE OF INVENTION: Streptavidin Mutins
; FILE REFERENCE: HUBR 119
; CURRENT APPLICATION NUMBER: US/08/948,097C
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: DE 196 41 876.3
; EARLIER FILING DATE: 1996-10-10
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: BINDING
; OTHER INFORMATION: Binding ligand for streptavidin
US-08-948-097-2

Query Match          93.8%; Score 45; DB 3; Length 8;
Best Local Similarity 75.0%; Pred. No. 4.1e+05;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 WWHQPER 8
       1:|||||:
       1 WSHQPER 8

Db

RESULT 2
US-09-382-950-7
; Sequence 7, Application US/09382950
; Patent No. 6303337
; GENERAL INFORMATION:
; APPLICANT: Rothchild, Kenneth
; APPLICANT: Gite, Sedand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
; FILE REFERENCE: AMBER-03879
; CURRENT APPLICATION NUMBER: US/09/382,950
; CURRENT FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent version 3.0
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial/Unknown
```

FEATURE:
NAME/KEY: misc feature
LOCATION: (1..7)
OTHER INFORMATION: Synthetic
US-09-382-950-7

Query Match 93.8%; Score 45; DB 3; Length 8;
Best Local Similarity 75.0%; Pred. No. 4.1e+05;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXPQFER 8
|:|||||:
Db 1 WSHPOFEK 8

RESULT 3
US-09-382-736B-8
Sequence 8, Application US/09382736B
Patent No. 6306628
GENERAL INFORMATION:

APPLICANT: Rothchild, Kenneth
APPLICANT: Gite, Sadanand
APPLICANT: Olejnik, Jerzy
TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Pro
FILE REFERENCE: AMBR-03951
CURRENT APPLICATION NUMBER: US/09/382,736B
CURRENT FILING DATE: 1999-08-25
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin version 3.0
SEQ ID NO 8
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-09-382-736B-8

Query Match 93.8%; Score 45; DB 3; Length 8;
Best Local Similarity 75.0%; Pred. No. 4.1e+05;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXPQFER 8
|:|||||:
Db 1 WSHPOFEK 8

RESULT 4
US-09-619-103-9
Sequence 9, Application US/09619103
Patent No. 6429300
GENERAL INFORMATION:

APPLICANT: Kurtz, Markus
APPLICANT: Lohse, Peter
APPLICANT: Wagner, Richard
TITLE OF INVENTION: Peptide Acceptor Ligation Methods
FILE REFERENCE: 50036/031002
CURRENT APPLICATION NUMBER: US/09/619,103
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 60/145,834
PRIOR FILING DATE: 1999-07-27
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: designed sequence to act as an identifying tag
US-09-619-103-9

Query Match 93.8%; Score 45; DB 4; Length 8;
Best Local Similarity 75.0%; Pred. No. 4.1e+05;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXPQFER 8
|:|||||:
Db 1 WSHPOFEK 8

RESULT 5
US-10-104-218-5
Sequence 5, Application US/10104218
Patent No. 6579705
GENERAL INFORMATION:

APPLICANT: MAIER, Thomas
APPLICANT: GABBERT, Carsten
TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
FILE REFERENCE: MAIER, T. ET AL.-2
CURRENT APPLICATION NUMBER: US/10/104,218
CURRENT FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: German No. 6579705 101 16 881.0
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: German No. 6579705 101 21 515.0
PRIOR FILING DATE: 2001-05-03
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patentin version 3.0
SEQ ID NO 5
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)..(8)
OTHER INFORMATION: StreptagII affinity peptide for protein purification
US-10-104-218-5

Query Match 93.8%; Score 45; DB 4; Length 8;
Best Local Similarity 75.0%; Pred. No. 4.1e+05;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 WXPQFER 8
|:|||||:
Db 1 WSHPOFEK 8

RESULT 6
US-09-809-517A-9
Sequence 9, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:

APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: Patentin version 3.0
SEQ ID NO 9
LENGTH: 8
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-9

Query Match 93.8%; Score 45; DB 4; Length 8;
Best Local Similarity 75.0%; Pred. No. 4.1e+05;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Db 1 WSHPOFEK 8

RESULT 7

US-09-809-517A-6

; Sequence 6, Application US/09809517A

; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

; TITLE OF INVENTION: Particles via disulfide bonds

; FILE REFERENCE: MORPHO/11

; CURRENT APPLICATION NUMBER: US/09/809,517A

; CURRENT FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: EP 99114072.4

; PRIOR FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: EP 00103551.8

; PRIOR FILING DATE: 2000-02-18

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 6

; LENGTH: 10

; TYPE: PRT

; ORGANISM: artificial sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-6

Query Match

Best Local Similarity 93.8%; Score 45; DB 4; Length 10;

Best Local Similarity 75.0%; Pred. No. 0.051;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 WSHPOFEK 8

Db 3 WSHPOFEK 10

RESULT 8

US-09-809-517A-30

; Sequence 30, Application US/09809517A

; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

; TITLE OF INVENTION: Particles via disulfide bonds

; FILE REFERENCE: MORPHO/11

; CURRENT APPLICATION NUMBER: US/09/809,517A

; CURRENT FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: EP 99114072.4

; PRIOR FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: EP 00103551.8

; PRIOR FILING DATE: 2000-02-18

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 30

; LENGTH: 21

; TYPE: PRT

; ORGANISM: artificial sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-30

Query Match

Best Local Similarity 93.8%; Score 45; DB 4; Length 21;

Best Local Similarity 75.0%; Pred. No. 0.11;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 WSHPOFEK 8

Db 14 WSHPOFEK 21

RESULT 9

US-09-809-517A-33

; Sequence 33, Application US/09809517A

; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

; TITLE OF INVENTION: Particles via disulfide bonds

; FILE REFERENCE: MORPHO/11

; CURRENT APPLICATION NUMBER: US/09/809,517A

; CURRENT FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: EP 99114072.4

; PRIOR FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: EP 00103551.8

; PRIOR FILING DATE: 2000-02-18

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 33

; LENGTH: 22

; TYPE: PRT

; ORGANISM: artificial sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-33

Query Match

Best Local Similarity 93.8%; Score 45; DB 4; Length 22;

Best Local Similarity 75.0%; Pred. No. 0.12;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 WSHPOFEK 8

Db 15 WSHPOFEK 22

RESULT 10

US-09-809-517A-31

; Sequence 31, Application US/09809517A

; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

; TITLE OF INVENTION: Particles via disulfide bonds

; FILE REFERENCE: MORPHO/11

; CURRENT APPLICATION NUMBER: US/09/809,517A

; CURRENT FILING DATE: 2001-03-15

; PRIOR APPLICATION NUMBER: EP 99114072.4

; PRIOR FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: EP 00103551.8

; PRIOR FILING DATE: 2000-02-18

; NUMBER OF SEQ ID NOS: 41

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 31

; LENGTH: 24

; TYPE: PRT

; ORGANISM: artificial sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-31

Query Match

Best Local Similarity 93.8%; Score 45; DB 4; Length 24;

Best Local Similarity 75.0%; Pred. No. 0.13;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 WSHPOFEK 8

Db 17 WSHPOFEK 24

RESULT 11

US-09-809-517A-34

; Sequence 34, Application US/09809517A

; Patent No. 6753136

; GENERAL INFORMATION:

; APPLICANT: Lohning, Corinna

; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

```

; TITLE OF INVENTION: particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match          93.8%; Score 45; DB 4; Length 25;
Best Local Similarity 75.0%; Pred. No. 0.14;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 WXHPOFER 8
DB      18 WSHPOFEK 25

RESULT 12
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

Query Match          93.8%; Score 45; DB 4; Length 117;
Best Local Similarity 75.0%; Pred. No. 0.75;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 WXHPOFER 8
DB      110 WSHPOFEK 117

RESULT 13
US-09-977-137A-5
; Sequence 5, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5
```

```

; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5

Query Match          93.8%; Score 45; DB 4; Length 117;
Best Local Similarity 75.0%; Pred. No. 0.75;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 WXHPOFER 8
DB      110 WSHPOFEK 117

RESULT 14
US-09-977-137A-7
; Sequence 7, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-7

Query Match          93.8%; Score 45; DB 4; Length 117;
Best Local Similarity 75.0%; Pred. No. 0.75;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 WXHPOFER 8
DB      110 WSHPOFEK 117

RESULT 15
US-09-977-137A-8
; Sequence 8, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8
```

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8
Query Match 93.8%; Score 45; DB 4; Length 117;
Best Local Similarity 75.0%; Pred. No. 0.75;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 WXHRPOFER 8
|:|||||:
DB 110 WSHRPOFER 117
Search completed: March 2, 2005, 12:25:33
Job time : 10.1057 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using ew model

Run on: March 2, 2005, 12:20:44 ; Search time 6.11382 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ4
Perfect score: 44
Sequence: 1 kxhpgfgg 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	88.6	131	2	S66537 Ig heavy chain V r
2	39	88.6	224	2	AF3382 alpha/beta hydroly
3	39	88.6	225	2	B97580 hypothetical prote
4	39	88.6	225	2	AH2800 conserved hypothet
5	39	88.6	436	1	S08156 gene T protein - m
6	39	88.6	512	2	H86216 protein T23G18.16
7	36	81.8	379	2	S55900 DNA-like protein
8	36	81.8	379	2	T41633 psi protein - flss
9	36	81.8	535	2	B82358 alkaline serine pr
10	36	81.8	563	2	S70684 pyruvate decarboxy
11	36	81.8	564	1	S36363 pyruvate decarboxy
12	36	81.8	564	2	S50700 pyruvate decarboxy
13	35	79.5	288	2	T12462 hypothetical prote
14	35	79.5	434	2	S42676 KES1 protein - yea
15	35	79.5	748	2	AG2486 exodeoxyribonuclea
16	35	79.5	10797	2	T30192 probable peptide s
17	34	77.3	181	2	T07661 maturation protein
18	34	77.3	186	2	B86178 hypothetical prote
19	34	77.3	295	2	AG3129 transcription regu
20	34	77.3	295	2	B98158 asgt protein (Af23
21	34	77.3	347	2	T17645 hypothetical prote
22	34	77.3	358	2	T48090 hypothetical prote
23	34	77.3	391	2	G83807 hypothetical prote
24	34	77.3	399	2	B69256 conserved hypothet
25	34	77.3	433	2	AD3570 galactarate dehydr
26	34	77.3	452	2	F83587 inner membrane pro
27	34	77.3	510	2	C69753 conserved hypothet
28	34	77.3	524	2	E97697 probable D-galacta
29	34	77.3	524	2	AE2923 D-galactarate dehy

30	33	75.0	208	2	AC3091 hypothetical prote
31	33	75.0	224	2	G98195 hypothetical prote
32	33	75.0	247	2	H98115 arginine deiminase
33	33	75.0	260	2	AB2080 hypothetical prote
34	33	75.0	276	2	AC0248 hypothetical prote
35	33	75.0	296	2	B71369 probable sugar ABC
36	33	75.0	299	2	G82222 conserved hypothet
37	33	75.0	302	2	F86849 transcription regu
38	33	75.0	309	2	B87712 conserved hypothet
39	33	75.0	322	2	H64933 succinylglutamate
40	33	75.0	322	2	F85783 hypothetical prote
41	33	75.0	322	2	B90935 hypothetical prote
42	33	75.0	328	2	H81996 probable integral
43	33	75.0	328	2	E81225 hypothetical prote
44	33	75.0	341	2	C82161 hypothetical prote
45	33	75.0	348	2	G72207 riboflavin bi funct

ALIGNMENTS

RESULT 1

S66537
Ig heavy chain V region - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000
C:Accession: S66537
R:Tsitolis, G.; Haase, W.; Engel, A.; Michel, H.
Eur. J. Biochem. 231, 823-830, 1995
A>Title: Isolation and structural characterization of trimeric cyanobacterial photosyete
A:Reference number: S66536; MUID:95377218; PMID:7649183
A:Accession: S66537
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-131 <TS1>
A:Cross-references: EMBL:X88902; NID:g895869; PIDN:CAA61364.1; PID:g1103701
C:Superfamily: immunoglobulin V region; immunoglobulin homology
F:5-8/Domain: immunoglobulin homology <IMM>

Query Match 88.6%; Score 39; DB 2; Length 131;
Best Local Similarity 85.7%; Pred. No. 3.8;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORFG 8
DB 125 RHPORFG 131

RESULT 2
AF3382
alpha/beta hydrolyase [imported] - Brucella melitensis (strain 16M)

C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AF3382
R:Delvecchio, V.G.; Kaparatel, V.; Redkar, R.J.; Patra, G.; Muir, C.; Los, T.; Ivanova, .; Masur, M.; Goettman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A>Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3382
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-224 <KUR>
A:Cross-references: UNIPROT:Q8YGM1; UNIPROT:Q8G102; GB:AE008917; PIDN:AAU52225.1; PID:g1
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME11044
A:Map position: I

Query Match 88.6%; Score 39; DB 2; Length 224;
Best Local Similarity 85.7%; Pred. No. 6.6;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8
:|||||
Db 33 PHPQFG 39

RESULT 3

hypothetical protein AGR_C_3351 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C/Species: Agrobacterium tumefaciens
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
C/Accession: B97580
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirolo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Leppas, C.; Markelz, B.;
Science 294, 2223-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A/Reference number: A97359; MUID:21608551; PMID:11743194
A/Accession: B97580
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-225 <KUR>
A/Cross-references: UNIPROT:Q8UED4; GB:AE007869; PIDN:AAK87595.1; PID:G15156937; GSPDB:G
A/Genes: AGR_C_3351
A/Map position: circular chromosome

Query Match 88.6%; Score 39; DB 2; Length 225;
Best Local Similarity 85.7%; Pred. No. 6.7;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8
:|||||
Db 33 PHPQFG 39

RESULT 4

conserved hypothetical protein Atu1826 [imported] - Agrobacterium tumefaciens (strain C5
C/Species: Agrobacterium tumefaciens
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
C/Accession: A12800
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kulyavich, T.; Levy, R.; Li, M.; McClell
Science 294, 2317-2323, 2001
A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: A12800
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-225 <KUR>
A/Cross-references: UNIPROT:Q8UED4; GB:AE008698; PIDN:AA142822.1; PID:G17740269; GSPDB:G
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Genes: Atu1826
A/Map position: circular chromosome

Query Match 88.6%; Score 39; DB 2; Length 225;
Best Local Similarity 85.7%; Pred. No. 6.7;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8
:|||||
Db 33 PHPQFG 39

RESULT 5

gene T protein - mouse
C/Species: Mus musculus (house mouse)
C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: S08156

R/Herrmann, B.G.; Label, S.; Poustka, A.; King, T.R.; Lehrach, H.
Nature 343, 617-622, 1990
A/Title: Cloning of the T gene required in mesoderm formation in the mouse.

A/Reference number: S08156; MUID:90158787; PMID:2154694
A/Accession: S08156
A/Status: not compared with conceptual translation
A/Molecule type: mRNA
A/Residues: 1-436 <HER>
A/Cross-references: UNIPROT:P20293; GB:X51683; NID:955053; PIDN:CAA35985.1; PID:955054
C/Genetics:
A/Map position: 17
A/Superfamily: mouse gene T protein; T-box homology
C/Keywords: DNA binding; transcription regulation
P/51-227/Domain: T-box homology <TBX>

Query Match 88.6%; Score 39; DB 1; Length 436;
Best Local Similarity 85.7%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8
:|||||
Db 260 SHPQFG 266

RESULT 6

protein T23G18.16 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C/Accession: H86216
R/Theologis, A.; Eckert, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chun, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A/Reference number: A86141; MUID:21016719; PMID:11130712
A/Accession: H86216
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-512 <STO>
A/Cross-references: UNIPROT:Q9SGD0; GB:AE005172; NID:96579199; PIDN:AA18242.1; GSPDB:GN
A/Genes: T23G18.16
A/Map position: 1

Query Match 88.6%; Score 39; DB 2; Length 512;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8
:|||||
Db 112 PHPQFG 118

RESULT 7

DNAI-like protein homolog - fission yeast (Schizosaccharomyces pombe)
C/Species: Schizosaccharomyces pombe
C/Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
C/Accession: S55900
R/Fair, S.K.; Chon, S.K.; Yoo, H.S.
Biochim. Biophys. Acta 1262, 87-90, 1995
A/Title: A cDNA of Schizosaccharomyces pombe encoding a homologue of DnaI-like protein.
A/Reference number: S55900; MUID:95290501; PMID:7772606
A/Accession: S55900
A/Status: preliminary
A/Molecule type: mRNA

A:Residues: 1-379 <PAR>
 A:Cross-references: UNIPROT:Q09912; EMBL:L37553; NID:g576932; PIDN:AAA74732.1; PID:g9532
 C:Gene: psi
 C:Superfamily: heat shock protein dnaJ, dnaJ amino-terminal homology <DNJ>
 F:6-68/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 81.8%; Score 36; DB 2; Length 379;
 Best Local Similarity 62.5%; Pred. No. 41;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKHPOFG 8
 DB 187 RSHPSFG 194

RESULT 8

T41633
 psi protein - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C:Accession: T41633
 R:McDougal, R.M.; Rajandream, M.A.; Barrell, B.G.; Rameberger, U.; Pohl, T.
 submitted to the EMBL Data Library, August 1999
 A:Reference number: 222005
 A:Accession: T41633
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-379 <MCD>
 A:Cross-references: UNIPROT:Q09912; EMBL:AL109850; PIDN:CA52880.1; GSPDB:GN00068; SPDB:
 C:Experimental source: strain 972h-; cosmid c830
 C:Gene: SPDB:SPCC830.07c
 A:Map position: 3
 A:Superfamily: heat shock protein dnaJ, dnaJ amino-terminal homology
 F:6-68/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 81.8%; Score 36; DB 2; Length 379;
 Best Local Similarity 62.5%; Pred. No. 41;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKHPOFG 8
 DB 187 RSHPSFG 194

RESULT 9

B82358
 alkaline serine proteinase VC0157 [imported] - Vibrio cholerae (strain N16961 serogroup
 C:Species: Vibrio cholerae
 C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
 C:Accession: B82358
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gilm, M.L.; Dodson, R.J.;
 chardson, D.; Emolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, F
 I. R.R.; Melandera, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A:Reference number: A82035; MUID:20406833; PMID:10952301
 A:Accession: B82358
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-535 <HEI>
 A:Cross-references: UNIPROT:Q9KVI8; GB:AE004105; GB:AE003852; NID:g9654551; PIDN:AAF9333
 A:Experimental source: serogroup O1, strain N16961; biotype El Tor
 C:Gene: VC0157
 A:Map position: 1
 C:Superfamily: subtilisin; subtilisin homology

Query Match 81.8%; Score 36; DB 2; Length 535;
 Best Local Similarity 71.4%; Pred. No. 59;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORG 8
 DB 184 NHPERG 190

RESULT 10

S70684
 pyruvate decarboxylase (EC 4.1.1.1) - yeast (Kluyveromyces marxianus var. lactis)
 C:Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
 C:Date: 15-Feb-1997 #sequence_revision 13-Mar-1997 #text_change 12-Jul-2004
 C:Accession: S70684
 R: Bianchi, M.M.; Tizani, L.; Destrelle, M.; Frontali, L.; Mesolowski-Louvel, M.
 Mol. Microbiol. 19, 27-36, 1996
 A:Title: The 'petite-negative' yeast Kluyveromyces lactis has a single gene expressing p
 A:Reference number: S70684; MUID:96419161; PMID:8821934
 A:Accession: S70684
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-563 <BIA>
 A:Cross-references: UNIPROT:Q12629; EMBL:X85968; NID:g1226006; PIDN:CA59953.1; PID:g122
 A:Note: the source is designated as Kluyveromyces lactis
 C:Gene: PDCL
 C:Superfamily: pyruvate decarboxylase; thiamin pyrophosphate-binding domain homology
 C:Keywords: carbon-carbon lyase; carboxy-lyase
 F:433-479/Domain: thiamin pyrophosphate-binding domain homology <TPB>

Query Match 81.8%; Score 36; DB 2; Length 563;
 Best Local Similarity 62.5%; Pred. No. 62;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKHPOFG 8
 DB 254 EKHPRFG 261

RESULT 11

S36363
 pyruvate decarboxylase (EC 4.1.1.1) - yeast (Kluyveromyces marxianus)
 N:Alternate names: alpha-ketocid carboxylase
 C:Species: Kluyveromyces marxianus
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Aug-2004
 C:Accession: S36363
 R:Holloway, P.; Suden, R.E.
 Curr. Genet. 24, 274-277, 1993
 A:Title: The isolation and nucleotide sequence of the pyruvate decarboxylase gene from K
 A:Reference number: S36363; MUID:94037155; PMID:8221939
 A:Accession: S36363
 A:Molecule type: DNA
 A:Residues: 1-564 <HOL>
 A:Cross-references: UNIPROT:P33149; EMBL:U09727; NID:g173308; PIDN:AAA35267.1; PID:g1733
 C:Gene: PDCL
 C:Superfamily: Pyruvate decarboxylase / indolepyruvate decarboxylase; thiamin pyrophosph
 C:Keywords: carbon-carbon lyase; carboxy-lyase
 F:433-479/Domain: thiamin pyrophosphate-binding domain homology <TPB>

Query Match 81.8%; Score 36; DB 1; Length 564;
 Best Local Similarity 62.5%; Pred. No. 62;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKHPOFG 8
 DB 254 EKHPRFG 261

RESULT 12

S50700
 pyruvate decarboxylase (EC 4.1.1.1) - yeast (Hanseniaspora uvarum)
 C:Species: Hanseniaspora uvarum
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 12-Jul-2004
 C:Accession: S50700
 R:Holloway, P.; Suden, R.E.

```

Yeast 10, 1581-1589, 1994
A>Title: The nucleotide sequence and initial characterization of pyruvate decarboxylase
A/Reference number: S50700; MUID:95242833; PMID:7725793
A/Accession: S50700
A/Molecule type: DNA
A/Residues: 1-564 <HOU>
A/Cross-references: UNIPROT:P34734; EMBL:U13635; NID:9535343; PID:AAA85103.1; PID:95353
A/Genes: PDC
C/Superfamily: pyruvate decarboxylase; thiamin pyrophosphate-binding domain homology
C/Keywords: carbon-carbon lyase; carboxy-lyase; homotetramer
F/433-479/Domain: thiamin pyrophosphate-binding domain homology <TPB>

Query Match
Best Local Similarity 81.8%; Score 36; DB 2; Length 564;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFG 8
DB 254 EKHPRFG 261

RESULT 13
T12462
Hypothetical protein DKFZP564I122.1 - human (fragment)
C/Species: Homo sapiens (man)
C/Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
C/Accession: T12462
R/Duesterhoeft A.; Lauber, J.; Mewes, H.W.; Gaassenhuber, U.; Wilmann, S.
submitted to the Protein Sequence Database, June 1999
A/Reference number: Z17522
A/Accession: T12462
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-288 <DUE>
A/Cross-references: UNIPROT:Q9YU1; EMBL:AL080062
A/Experimental source: fetal brain; clone DKFZP564I122
A/Note: DKFZP564I122.1

Query Match
Best Local Similarity 79.5%; Score 35; DB 2; Length 288;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
DB 156 IHRPFG 162

RESULT 14
S42676
KE81 protein - yeast (Saccharomyces cerevisiae)
N/Alternate names: protein P2614; protein YPL145C
C/Species: Saccharomyces cerevisiae
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C/Accession: S42676; S65156; S69042; S69454
R/Jiang, B.; Brown, J.L.; Sheraton, J.; Fortin, N.; Bussey, H.
Yeast 10, 341-353, 1994
A/Title: A new family of yeast genes implicated in ergosterol synthesis is related to th
A/Reference number: S42676; MUID:94287711; PMID:8017104
A/Accession: S42676
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-434 <JIA>
A/Cross-references: UNIPROT:P35844; GB:U03913; NID:9433362; PID:AAA1736.1; PID:9433363
R/Punelle, B.; Coster, F.; Goffeau, A.
submitted to the Protein Sequence Database, May 1996
A/Reference number: S65154
A/Accession: S65156
A/Molecule type: DNA
A/Residues: 1-434 <PUR>
A/Cross-references: EMBL:Z73501; NID:91370309; PID:e247042; PID:91370310; MIPS:YPL145C
A/Experimental source: strain S288C (AB972)

```

```

R.Hall, J.; Depaulo, T.; Ahmed, A.; Bussey, H.; Fortin, N.; Friesen, J.D.; Storms, R.K.
submitted to the EMBL Data Library, December 1995
A/Description: The sequence of Saccharomyces cerevisiae chromosome XVI left arm.
A/Reference number: S69040
A/Accession: S69042
A/Molecule type: DNA
A/Residues: 1-434 <HAL>
A/Cross-references: EMBL:U43703; NID:91244769; PID:91244772; MIPS:YPL145C
R/Punelle, B.; Comblez, S.; Coster, F.; Naveau, F.; Goffeau, A.
submitted to the EMBL Data Library, March 1996
A/Description: The sequence of 35 kb on the left arm of yeast chromosome XVI identifies
a gene to the human phosphotyrosyl phosphatase activator PTPA and a homologue to the plant
A/Reference number: S69428
A/Accession: S69454
A/Molecule type: DNA
A/Residues: 1-434 <PUW>
A/Cross-references: EMBL:X96770; NID:91403537; PID:e239051; PID:91403564
A/Genes: SGD:KES1
A/Genes: SGD:KES1
A/Map position: 16L

Query Match
Best Local Similarity 79.5%; Score 35; DB 2; Length 434;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFG 7
DB 126 KEMPRFG 132

RESULT 15
AG2486
exodeoxyribonuclease V, alpha chain [imported] - Nostoc sp. (strain PCC 7120) plasmid pc
C/Species: Nostoc sp. PCC 7120
A/Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C/Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C/Accession: AG2486
R/Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A/Reference number: AB1807; MUID:21595285; PMID:11759840
A/Accession: AG2486
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-748 <KUR>
A/Cross-references: UNIPROT:Q8YL64; GB:BA000020; PID:BA078155.1; PID:917135609; GSPDB:G
C/Genes:
A/Genes: all7071
A/Genome: plasmid

Query Match
Best Local Similarity 79.5%; Score 35; DB 2; Length 748;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFG 7
DB 71 KXHPQFG 77

Search completed: March 2, 2005, 12:28:50
Job time : 7.11382 secs

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 29.7236 Seconds
(without alignments)
137.824 Million cell updates/sec

Title: SEQ4
Perfect score: 44
Sequence: 1 kxhpgfgs 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Uniprot_03:*
1: uniprot_sprot:*
2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	88.6	67	2	Q8S212
2	39	88.6	92	2	Q02737
3	39	88.6	212	2	Q97669
4	39	88.6	224	2	Q87GM1
5	39	88.6	224	2	Q8G102
6	39	88.6	225	2	Q92PK7
7	39	88.6	225	2	Q8UED4
8	39	88.6	226	2	Q6FZM3
9	39	88.6	226	2	Q6G3B7
10	39	88.6	228	2	Q98NR9
11	39	88.6	260	2	Q750S3
12	39	88.6	330	2	Q96UF7
13	39	88.6	385	2	Q96UF9
14	39	88.6	385	2	Q9P3Y1
15	39	88.6	386	2	Q14410
16	39	88.6	435	1	BRAC_HUMAN
17	39	88.6	435	1	Q9GL27
18	39	88.6	436	1	BRAC_MOUSE
19	39	88.6	512	2	Q9SGD0
20	39	88.6	1074	2	Q9W3W1
21	38	86.4	252	2	Q7Y3G7
22	38	86.4	525	2	Q8G3H6
23	38	86.4	780	2	P77846
24	38	86.4	2383	2	Q840C8
25	36	81.8	232	2	Q7JH66
26	36	81.8	282	1	MIOX_PIG
27	36	81.8	294	2	Q7UVW5
28	36	81.8	309	1	KFRS_WOLSU
29	36	81.8	323	2	Q6TR93
30	36	81.8	356	2	Q8CNS6
31	36	81.8	379	1	PS11_SCHPO

32	36	81.8	390	2	Q96UF8	Q96UF8 metarhizium
33	36	81.8	390	2	Q9UR5	Q9UR5 metarhizium
34	36	81.8	428	2	Q6S912	Q6S912 hansenula a
35	36	81.8	440	2	Q6C012	Q6C012 yarrowia il
36	36	81.8	465	2	Q8SQJ3	Q8SQJ3 encephalito
37	36	81.8	535	2	Q9KVI8	Q9KVI8 vibrio chol
38	36	81.8	563	1	DCPY_KLUUA	Q12629 kluyveromyc
39	36	81.8	563	2	Q8J134	Q8J134 candida gla
40	36	81.8	563	2	Q75F01	Q75F01 ashbya gos
41	36	81.8	564	1	DCPY_HANUV	P34734 hanseniaspo
42	36	81.8	564	1	DCPY_KLUUA	P33149 kluyveromyc
43	36	81.8	564	2	Q6FUJ3	Q6FUJ3 candida gla
44	36	81.8	564	2	Q6W9G6	Q6W9G6 saccharomyc
45	36	81.8	564	2	Q6XAE9	Q6XAE9 saccharomyc

ALIGNMENTS

```

RESULT 1
ID Q8S212 PRELIMINARY; PRT; 67 AA.
AC Q8S212;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
DE LD08931P.
GN ORFNames=CG4532;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Berkely;
RA Stapleton M., Brokslein P., Hong L., Agbayani A., Carlson J.,
RA Champagne M., Chavez C., Dorsett V., Dreenek D., Farfan D., Frise E.,
RA George R., Gonzalez M., Guarin H., Krommler B., Li P., Liao G.,
RA Miranda A., Mungali C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA Celniker S.;
RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AY070876; AA148498.1; -.
DR FlyBase; FBgn029903; CG4532.
SQ SEQUENCE 67 AA; 7782 MW; F9C917B81577B94 CRC64;

Query Match 88.6%; Score 39; DB 2; Length 67;
Best Local Similarity 85.7%; Pred. No. 6.4;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 19 DHPQFG 25

RESULT 2
ID Q02737 PRELIMINARY; PRT; 92 AA.
AC Q02737;
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Brachyury (Fragment).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxId=9823;
RN (1)
RP SEQUENCE FROM N.A.
RA Yelich J.V., Pomp D., Geisert R.D.;
RL Submitted (FEB-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL; U91519; AAB53317.1; -.

```

DR GO: GO:0005634; C:nucleus; IEA.
 DR GO: GO:0003700; F:transcription factor activity; IEA.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR002070; TF Brachyury.
 DR PRINTS: PR00938; BRACHYURY.
 FT NON_TER 1
 FT NON_TER 92
 FT NON_TER 92
 SQ SEQUENCE 92 AA; 10134 MW; 7AE235EB7B32AD7D CRC64;

Query Match
 Best Local Similarity 88.6%; Score 39; DB 2; Length 92;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 XHPQFG 8
 Db 51 PHPQFG 57

RESULT 3

097669 PRELIMINARY; PRT; 212 AA.
 AC 097669;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DE Brachyury (T) protein (Fragment).
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 NCBI_TaxId=9986;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=New Zealand White;
 RX MEDLINE=2196934; PubMed=11973268;
 RA Viebahn C., Storz C., Mitchell S.A., Blum M.;
 RT "Low proliferative and high migratory activity in the area of
 Brachyury expressing mesoderm progenitor cells in the gastrulating
 rabbit embryo.";
 RL Development 129:2355-2365(2002).
 DR EMBL: AF102131; AAD11790.1; --
 DR HSSP; P24781; IXHR
 DR GO: GO:0005634; C:nucleus; IEA.
 DR GO: GO:0003700; F:transcription factor activity; IEA.
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro: IPR008967; P53 like DNA_bnd.
 DR InterPro: IPR002070; TF Brachyury.
 DR Pfam: PF00907; T-box; 1
 DR PRINTS: PR00938; BRACHYURY.
 DR SMART: SM00425; TBOX; 1.
 DR PROSITE: PS00252; TBOX_3; 1.
 FT NON_TER 1
 FT NON_TER 212
 FT NON_TER 212
 SQ SEQUENCE 212 AA; 22884 MW; CB7FD39EA372B4C CRC64;

Query Match
 Best Local Similarity 88.6%; Score 39; DB 2; Length 212;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 XHPQFG 8
 Db 114 PHPQFG 120

RESULT 4

08YGM1 PRELIMINARY; PRT; 224 AA.
 AC 08YGM1;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
 DE ALPHA/BETA HYDROLASE.

GN OrderedLocNames=EMBL1044;
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 NCBI_TaxId=29459;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / Biotype 1;
 RX MEDLINE=20020109; PubMed=11756688; DOI=10.1073/pnas.221575398;
 RA DelVecchio V.G., Kaprielian V., Redkar R.J., Patra G., Mujar C., Los T.,
 RA Ivanova N., Anderson I., Bhattacharya A., Lykide A., Reznik G.,
 RA Jblonk E., Elzer P.H., Hagius S., O'Callaghan D., Leeseon J.,
 RA Salkov R., Kyrpides N.C., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 Brucella melitensis.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL: AE009544; AA15225.1; --
 DR PIR: AF3382; AF3382.
 DR GO: GO:0003824; F:catalytic activity; IEA.
 KW Complete proteome.

Query Match
 Best Local Similarity 88.6%; Score 39; DB 2; Length 224;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 XHPQFG 8
 Db 33 PHPQFG 39

RESULT 5

08G102 PRELIMINARY; PRT; 224 AA.
 AC 08G102;
 DT 01-MAR-2003 (TREMBlrel. 23, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DE Hypothetical protein.
 GN OrderedLocNames=BR0929;
 OS Brucella suis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 NCBI_TaxId=29461;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=1330 / Biovar 1;
 RX MEDLINE=22247741; PubMed=12271122; DOI=10.1073/pnas.192319099;
 RA Paulsen I.T., Seeshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
 RA Read T.D., Dodson R.J., Umayam L.A., Brinkac L.M., Beaman M.J.,
 RA Naugherthy S.C., Debey R.T., Durkin A.S., Kolonay J.F., Madupu R.,
 RA Nelson W.C., Ayodeji B., Kral M., Shetty J., Malek J.A.,
 RA Van Aken S.E., Riedmuller S., Tettelin H., Gill S.R., White O.,
 RA Salzberg S.L., Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M.,
 RA Fraser C.M.;
 RT "The Brucella suis genome reveals fundamental similarities between
 animal and plant pathogens and symbionts.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
 DR EMBL: AE014395; AA029855.1; --
 DR PIR: AF3382; AF3382.
 DR TIGR: BR0929;
 DR GO: GO:0003824; F:catalytic activity; IEA.
 DR InterPro: IPR000379; Ser_esters.
 KW Complete proteome.

Query Match
 Best Local Similarity 88.6%; Score 39; DB 2; Length 224;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 2 XHPQFG 8

```

Db          33 PHPOFG 39

RESULT 6
Q92PK7      PRELIMINARY;      PRT;      225 AA.
ID 092PK7;
AC 092PK7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein SMC00528.
GN ORFNames=SMC00528;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxId=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;
RA Boland P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godle T., Goffeau A., Kahn D., Kiss B., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rampsberger U.,
RA Renard C., Thebaud P., Vandenberg M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591788; CAC46316.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR000379; Ser esters.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 225 AA; 25011 MW; C87CD753E13D92B CRC64;

Query Match          88.6%; Score 39; DB 2; Length 225;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORFG 8
Db          33 PHPOFG 39

RESULT 7
Q8UED4      PRELIMINARY;      PRT;      225 AA.
ID Q8UED4; Q7CYF9;
AC Q8UED4; Q7CYF9;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Hypothetical protein Atu1826 (AGR_C_3351p).
GN OrderedLocNames=AGR_C_3351, Atu1826;
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxId=176299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Dupont;
RX MEDLINE=21608550; PubMed=11743193; DOI=10.1126/science.1066804;
RA Wood D.W., Seubel J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.B., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland G., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chunley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Neeter E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Cereon;
RX MEDLINE=21608551; PubMed=11743194; DOI=10.1126/science.1066803;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houtmel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Planagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cleio C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AE009137; AAL42822.1; -.
DR EMBL; AE008102; AAK87595.1; -.
DR FIR; AH2800; AH2800.
DR PIR; B97580; B97580.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR000379; Ser esters.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 225 AA; 25084 MW; 1B7BDF5600841833 CRC64;

Query Match          88.6%; Score 39; DB 2; Length 225;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORFG 8
Db          33 PHPOFG 39

RESULT 8
Q6FZW3      PRELIMINARY;      PRT;      226 AA.
ID Q6FZW3;
AC Q6FZW3;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocNames=B005930;
OS Bartonella quintana (Rochalimaea quintana).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxId=803;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Toulouse;
RX PubMed=15210978; DOI=10.1073/pnas.0305659101;
RA Almaraz U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardeli D.H.,
RA Canhaek B., Eriksson A.-S., Naeslund A.K., Handley S.A., Huver M.,
RA La Scola B., Holmberg M., Andersson S.G.E.;
RT "The louse-borne human pathogen Bartonella quintana is a genomic
RT derivative of the zoonotic agent Bartonella henselae."
RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).
DR EMBL; BX897700; CAF26085.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR000379; Ser esters.
KW Complete proteome.
SQ SEQUENCE 226 AA; 25373 MW; D2AD7C07EBC2B625 CRC64;

Query Match          88.6%; Score 39; DB 2; Length 226;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORFG 8
Db          33 PHPOFG 39

RESULT 9
Q6G3B7      PRELIMINARY;      PRT;      226 AA.
ID Q6G3B7;
AC Q6G3B7;

```

DT 05-JUL-2004 (TReMBLrel. 27, Created)
 DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
 DE Hypothetical protein.
 GN OrderedLocNames=BH08660;
 OS Bartonella henselae (Rochalimaea henselae).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bartonellaceae; Bartonella.
 RN NCBI_Taxid=38323;
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 49882 / Houston 1;
 RX PubMed=15210978; DOI=10.1073/pnas.030569101;
 RA Altmann U.C.M., Frank A.C., Karberg E.O., Legault B.-A., Ardehl D.H.,
 RA Canbacek B., Eriksson A.-S., Naslund A.K., Handley S.A., Huvel M.,
 RA La Scola B., Holmberg M., Andersson S.G.E.;
 RT "The house-borne human pathogen Bartonella quintana is a genomic
 RL derivative of the zoonotic agent Bartonella henselae.";
 DR EMBL BX897699; CAF27664.1; -
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR InterPro; IPR000379; Ser ester.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 226 AA; 25230 MW; 7216738444690A60 CRC64;

Query Match
 Best Local Similarity 88.6%; Score 39; DB 2; Length 226;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8
 DB 33 PHPQFG 39

RESULT 10
 ID Q96NR9 PRELIMINARY; PRT; 228 AA.
 AC Q96NR9;
 DT 01-OCT-2001 (TReMBLrel. 18, Created)
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
 DE M110014 protein.
 GN OrderedLocNames=m110014;
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Mesorhizobium.
 RN NCBI_Taxid=381;
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF30309;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimoto M., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RL Mesorhizobium loti.";
 DR EMBL F7331-338(2000).
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR InterPro; IPR000379; Ser ester.
 KW Complete proteome.
 SQ SEQUENCE 228 AA; 25544 MW; B1A116CBEA893BC CRC64;

Query Match
 Best Local Similarity 88.6%; Score 39; DB 2; Length 228;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8
 DB 37 PHPQFG 43

RESULT 11
 ID Q750S3 PRELIMINARY; PRT; 260 AA.
 AC Q750S3;
 DT 05-JUL-2004 (TReMBLrel. 27, Created)
 DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
 DE AG1134CP.
 GN ORFNames=AG1134C;
 OS Ashybya gossypii (Yeast) (Eremothecium gossypii).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
 RN NCBI_Taxid=33169;
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 10895;
 RA Lerch A., Brachet S., Voegel S.E., Gaffney T., Philippson P.,
 RA Dierich F.S.;
 RT Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL AF016820; AAS54357.1; -
 DR AGD; AG1134C; -
 DR GO; GO:0009982; F:pseudouridine synthase activity; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0006396; P:RNA processing; IEA.
 DR InterPro; IPR006145; Pseudou synth.
 DR Pfam; PF00849; Pseudou synth_2; 1.
 SQ SEQUENCE 260 AA; 29476 MW; EBF0765D04A8F447 CRC64;

Query Match
 Best Local Similarity 88.6%; Score 39; DB 2; Length 260;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KXHPFG 8
 DB 51 KXHPFG 58

RESULT 12
 ID Q96UF7 PRELIMINARY; PRT; 330 AA.
 AC Q96UF7;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DE 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
 GN Subtilisin-like serine protease PR1B (Fragment).
 OS Name=PR1B;
 OS Metarhizium anisopliae var. acridum.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Hypocreomycetidae; Hypocreales; Clavicipitaceae;
 OC Clavicipitaceae; Metarhizium.
 RN NCBI_Taxid=92637;
 RP SEQUENCE FROM N.A.
 RC STRAIN=FI-985;
 RA Baggia S., St Leger R.J.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL AJ416692; CAC95046.1; -
 DR HSSP; P06873; 2PRK
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilisin activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000209; Pept_S8_S53.
 DR InterPro; IPR009020; Prot_inh_propept.
 DR InterPro; IPR010259; Prot_inh_58A.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF05922; Subtilisin_N; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASIN_ASP; 1.
 KW Protease.
 FT NON_TER 330 330
 SQ SEQUENCE 330 AA; 34673 MW; 09B294C2864E9130 CRC64;

```

Query Match      88.6%; Score 39; DB 2; Length 330;
Best Local Similarity 85.7%; Pred. No. 35;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
      :|||||
Db      149 THPQFG 155

RESULT 13
Q96UF9 PRELIMINARY; PRT; 385 AA.
AC Q96UF9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DE Subtilisin-like protease PRIB.
GN Name=PRIB;
OS Metarhizium anisopliae var. anisopliae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Clavicipitaceae;
OC mitosporic clavicipitaceae; Metarhizium.
OX NCBI_TaxId=92636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARSEF 820;
RA Baga S., St Leger R.J.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ416690; CAC95044.1; -.
DR HSP; P06873; IIC6.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000183; Decarboxylase2.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR010259; Prot_inh_S8A.
DR Pfam; PF00082; Peptidase_S8; I.
DR PRINTS; PR00723; SUBTILISIN_N; 1.
DR PROSITE; PS00879; ODR_DC_2_2; UNKNOWN_1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
KW Protease.
SQ SEQUENCE 385 AA; 40094 MW; DB31BA0DC3A29FC3 CRC64;

Query Match      88.6%; Score 39; DB 2; Length 385;
Best Local Similarity 85.7%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
      :|||||
Db      149 THPQFG 155

RESULT 14
Q9P3Y1 PRELIMINARY; PRT; 385 AA.
AC Q9P3Y1;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Subtilisin-like protease PRIB.
GN Name=PRIB;
OS Metarhizium anisopliae var. anisopliae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Clavicipitaceae;
OC mitosporic clavicipitaceae; Metarhizium.
OX NCBI_TaxId=92636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARSEF 2575;
RA Baga S., Screen S.E., St Leger R.J.;
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.

```

```

DR EMBL; AJ289823; CAB95012.1; -.
DR HSP; P06873; IIC6.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000183; Decarboxylase2.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR009020; Prot_inh_propept.
DR InterPro; IPR010259; Prot_inh_S8A.
DR Pfam; PF00082; Peptidase_S8; I.
DR Pfam; PF05922; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00879; ODR_DC_2_2; UNKNOWN_1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
KW Protease.
SQ SEQUENCE 385 AA; 39959 MW; D9857DEFEC25D7C CRC64;

Query Match      88.6%; Score 39; DB 2; Length 385;
Best Local Similarity 85.7%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
      :|||||
Db      149 THPQFG 155

RESULT 15
Q14410 PRELIMINARY; PRT; 386 AA.
ID Q14410;
AC Q14410;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Subtilisin-like protease PRIB (Fragment).
GN Name=PRIB;
OS Metarhizium anisopliae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Clavicipitaceae;
OC mitosporic clavicipitaceae; Metarhizium.
OX NCBI_TaxId=5530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ARSEF 2575;
RX MEDLINE=97473490; PubMed=9332344; DOI=10.1016/S0378-1119(97)00132-7;
RA Joshi L., St Leger R.J., Roberts D.W.;
RT "Isolation of a cDNA encoding a novel subtilisin-like protease (PrIB)
RT from the entomopathogenic fungus, Metarhizium anisopliae using
RT differential display-RT-PCR."
RL Gene 197:1-8(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ARSEF 2575;
RA Joshi L.S.T., Leger R.J., Roberts D.W.;
RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL; U59484; AAC49831.1; -.
DR HSP; P06873; 2PRK.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000183; Decarboxylase2.
DR InterPro; IPR000209; Pept_S8_S53.
DR InterPro; IPR009020; Prot_inh_propept.
DR InterPro; IPR010259; Prot_inh_S8A.
DR Pfam; PF00082; Peptidase_S8; I.
DR Pfam; PF05922; Subtilisin_N; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00879; ODR_DC_2_2; UNKNOWN_1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
KW Protease.
FT NON_TER 386
SQ SEQUENCE 386 AA; 40031 MW; E805E0751C471568 CRC64;

Query Match      88.6%; Score 39; DB 2; Length 386;

```

Wed Mar 2 14:27:11 2005

seq4.rup

Page 6

Best Local Similarity 85.7%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches

Qy 2 XHPQFGS 8

0; Indels 0; Gaps 0;

Db 150 THPQFGS 156

Search completed: March 2, 2005, 12:44:13
Job time : 31.7236 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 / Search time 36.0325 Seconds
(without alignments)
85.869 Million cell updates/sec

Title: SEQ4
Perfect score: 44
Sequence: 1 kxhpqfgg 8

Scoring table: BLOSUM62DX
Gapop 10.0, Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: A_Geneseq_16dec04:*

- 1: geneseqp19808:*
- 2: geneseqp19908:*
- 3: geneseqp20008:*
- 4: geneseqp20018:*
- 5: geneseqp20028:*
- 6: geneseqp20038:*
- 7: geneseqp20038:*
- 8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	88.6	8	ABP60360	Abp60360 Streptavi
2	39	88.6	9	AAW32485	Aaw32485 Strept-tag
3	39	88.6	9	AAW59211	Aaw59211 Streptavi
4	39	88.6	9	AAW44010	Aaw44010 Strept pep
5	39	88.6	9	AAW06914	Aaw06914 Strept-tag
6	39	88.6	9	AAW30795	Aaw30795 Amino aci
7	39	88.6	9	ABG31053	Abg31053 Peptide p
8	39	88.6	9	AAW78476	Aaw78476 Strept-tag
9	39	88.6	9	AAW16086	Aaw16086 Neurologi
10	39	88.6	9	ABP55547	Abp55547 Streptag
11	39	88.6	9	AAE37229	Aae37229 Strept-tag
12	39	88.6	9	AAE32860	Aae32860 Streptag
13	39	88.6	9	AAE33270	Aae33270 Strept epi
14	39	88.6	9	ABG72479	Abg72479 Strept tag
15	39	88.6	9	ADB84587	Adb84587 Streptavi
16	39	88.6	9	ADJ14695	Adj14695 Strept pep
17	39	88.6	9	ADJ36219	Adj36219 Self-coal
18	39	88.6	9	ADJ88138	Adj88138 Strept tag
19	39	88.6	9	ADN29604	Adn29604 Strept tag
20	39	88.6	9	ADQ82347	Adq82347 Strept tag
21	39	88.6	10	AAW86446	Aaw86446 Streptavi
22	39	88.6	10	AAW50025	Aaw50025 Peptide t
23	39	88.6	10	AAW57325	Aaw57325 Human IL-
24	39	88.6	10	AAW67424	Aaw67424 Streptavi
25	39	88.6	10	AAW19071	Aaw19071 Peptide b

26	39	88.6	10	4	AAW61168	Aaw61168 Paramagne
27	39	88.6	10	4	AAW99026	Aaw99026 Streptavi
28	39	88.6	10	4	AAW02084	Aaw02084 Streptavi
29	39	88.6	10	5	AAW14759	Aaw14759 Human int
30	39	88.6	10	7	ADD29929	Add29929 Antibody-
31	39	88.6	11	2	AAW52691	Aaw52691 PASK46-p1
32	39	88.6	11	2	AAW52692	Aaw52692 PASK46-p1
33	39	88.6	12	2	AAW52696	Aaw52696 PASK46-p1
34	39	88.6	12	2	AAW81821	Aaw81821 Streptavi
35	39	88.6	18	5	AAW78473	Aaw78473 linker co
36	39	88.6	23	5	AAW37002	Aaw37002 Artificial
37	39	88.6	23	5	AAW19984	Aaw19984 Peptide o
38	39	88.6	42	5	AAW19988	Aaw19988 Protein o
39	39	88.6	42	5	AAW19986	Aaw19986 Protein o
40	39	88.6	46	8	AAW32686	Aaw32686 Receptor
41	39	88.6	46	8	AAW55658	Aaw55658 ED peptid
42	39	88.6	47	2	AAW51892	Aaw51892 VHL leade
43	39	88.6	47	5	AAW81776	Aaw81776 Plasmid p
44	39	88.6	132	8	ADJ36263	Adj36263 Self-coal
45	39	88.6	155	2	AAW32483	Aaw32483 Kappa 119

ALIGNMENTS

RESULT 1
ABP60360 standard; peptide; 8 AA.

XX	ABP60360;	
XX	28-MAR-2003 (first entry)	
XX	Streptavidin tag peptide SEQ ID NO 1.	
DE	Streptavidin; protein chip; microtitre plate; detection.	
KW	Synthetic.	
OS	DE10113776-A1.	
PN	02-OCT-2002.	
PD	21-MAR-2001; 2001DE-01013776.	
XX	21-MAR-2001; 2001DE-01013776.	
PF	21-MAR-2001; 2001DE-01013776.	
XX	21-MAR-2001; 2001DE-01013776.	
XX	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.	
PA	Schmidt T;	
XX	WPI; 2003-031166/03.	
DR	New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.	
PT	Disclosure; Page 2; 18pp; German.	
XX	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it may not be essential to remove it); facilitates detection and has easily controllable binding properties. (I) is particularly used for purifying FP from dilute solution in batch formats (which use simpler apparatus than column methods and result in lower loss of FP). The present sequence is that of a streptavidin tag peptide disclosed with the invention	

XX Sequence 8 AA;
SQ

Query Match 88.6%; Score 39; DB 6; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
:|||||
DB 2 RHPQFG 8

RESULT 2
AAW32485

ID AAW32485 standard; peptide; 9 AA.

AC AAW32485;

XX 27-MAR-1998 (first entry)

DE Strep-tag peptide.

KM Catalytic antibody; growth factor; B-cell mitogenesis; LHL; L protein;

XX hen egg lysozyme; strep-tag; purification.

OS Synthetic.

PN WO9735887-A1.

XX 02-OCT-1997.

PF 26-MAR-1997; 97WO-AU000194.

XX 26-MAR-1996; 96AU-00008951.

PR 27-FEB-1997; 97AU-00005375.

XX (AMRA-) AMRAD OPERATIONS PTY LTD.

PI Koentgen F, Sues GM, Tarlinton DM, Treutlein HR;

DR WPI; 1997-489572/45.

PT New catalytic antibody precursors - comprising a B-cell surface molecule
binding portion which can induce B-cell mitogenesis.

PS Example 3; Page 40; 109pp; English.

CC This peptide comprises the strep-tag peptide that is recognised by
CC streptavidin. A form of novel growth factor LHL (see AAW32479) was
CC generated by PCR that contains a FLAG epitope (see AAW32484) at its N-
CC terminus and the strep-tag at its C-terminus. The construct is designated
CC LHL-seq (see AAW32481), where L is the immunoglobulin binding entity from
CC *Streptococcus magnus* and H is residues 42-62 of hen egg lysozyme.
CC The strep-tag was used for purification of LHL-seq over a streptavidin
CC column. Thus, the LHL-seq was not purified on the basis of binding
CC immunoglobulin, thereby eliminating potential contamination by other
CC bacterial proteins which also bind immunoglobulins. LHL-seq has identical
CC activity to that of LHL and can be used in novel methods for the
CC generation of catalytic antibodies

XX Sequence 9 AA;

Query Match 88.6%; Score 39; DB 2; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
:|||||
DB 3 RHPQFG 9

RESULT 3
AAW59211

ID AAW59211 standard; peptide; 9 AA.

AC AAW59211;

XX 27-AUG-1998 (first entry)

DE Streptavidin tagged peptide ligand #1.

KM Streptavidin; ligand; binding affinity; mutant; isolation; purification;

XX recover; immobilise.

OS Synthetic.

PN EP835934-A2.

XX 15-APR-1998.

PF 09-OCT-1997; 97EP-00117504.

XX 10-OCT-1996; 96DE-01041876.

PA (BIOA-) INST BIOMALYTIK GMBH.

PI Skerra A, Voss S;

DR WPI; 1998-218868/20.

PT Streptavidin mutants with higher binding affinity for peptide ligands -
have mutation in amino acid region 44-53, used to isolate, purify or
determine fusion proteins including these ligands.

PS Claim 10; Page 11; 21pp; German.

CC AAW59211 and AAW59212 are ligands used in a method to assay binding
CC affinity of streptavidin mutants. These mutants have a mutation within
CC the amino acid (aa) region 44-53 of the wild-type protein show a higher
CC binding affinity than the wild-type for peptide ligands that include the
CC sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z
CC are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin
CC mutants can be used to isolate, purify and determine proteins or to
CC determine/recover substances that contain streptavidin-binding groups.
CC Such compounds may also be used to immobilise fusions on microtitre
CC plates, microbeads or sensor chips

XX Sequence 9 AA;

Query Match 88.6%; Score 39; DB 2; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
:|||||
DB 3 RHPQFG 9

RESULT 4
AAW44010

ID AAW44010 standard; peptide; 9 AA.

AC AAW44010;

XX 05-JUN-1998 (first entry)

DE Strep peptide epitope used in an epitope tagged prion protein construct.

KM Prion; epitope; FLAG; Strep; poly-histidine; haemagglutinin; recombinant;

XX transgenic animal; scrapie; Creutzfeldt-Jacob disease; CJD;

XX bovine spongiform encephalopathy; BSE.

OS Synthetic.
XX WO9746572-A1.


```

PD 11-DEC-1997.
XX
XX 29-MAY-1997; 97WO-US009289.
XX
XX 06-JUN-1996; 96US-0060626.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Prusiner SB, Telling GC, Cohen FE, Scott MR;
XX
XX WPI; 1998-042112/04.
XX
XX Nucleic acid construct encoding biologically active protein and epitope -
XX especially epitope-tagged prion protein.
XX
XX Claim 4; Page 48; 62pp; English.
XX
XX This sequence represents an artificial Strep peptide epitope. It is used
XX in a recombinant nucleic acid construct encoding an epitope-tagged prion
XX protein (Prp). The construct comprises a first nucleic acid sequence
XX encoding an amino acid sequence of a biologically active protein fragment
XX and a second nucleic acid sequence encoding a heterologous epitope
XX domain. The heterologous epitope domain is a peptide selected from a
XX peptide group of FLAG, Strep, poly-histidine, human c-myc peptide
XX recognised by monoclonal antibody 9E10 and haemagglutinin peptide
XX recognised by monoclonal antibody 12CA5. The protein is a natural,
XX synthetic or chimeric Prp molecule. The protein has two different, three-
XX dimensional conformations and the epitope domain is spatially positioned
XX relative to the protein such that the epitope domain is more exposed in a
XX first conformation relative to a second conformation. The nucleic acid
XX construct may be used for the production of transgenic animals or cells
XX that are useful in a method for distinguishing between different
XX conformational shapes of a protein. These methods are particularly useful
XX in studying diseases caused by prion proteins, e.g. Creutzfeldt-Jakob
XX disease (CJD), scrapie and bovine spongiform encephalopathy (BSE)
XX
XX Sequence 9 AA;
XX
XX Query Match 88.6%; Score 39; DB 2; Length 9;
XX Best Local Similarity 85.7%; Pred. No. 1.8e+06;
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 XHPOFGG 8
XX :|||||
XX 3 RHPQFGG 9
XX
XX Db
XX
XX RESULT 5
XX AAY06914
XX ID AAY06914 standard; peptide; 9 AA.
XX
XX AC AAY06914;
XX
XX DT 01-JUL-1999 (first entry)
XX
XX DE Strep-tag sequence.
XX
XX KW Growth factor precursor; B-cell surface; T cell surface; Cab; hepatitis;
XX catalytic antibody; immunoglobulin; tumour necrosis factor; influenza;
XX rheumatoid arthritis; toxic shock syndrome; viral docking receptor;
XX human immune virus; tumour-specific antigen; amyloid plaque; myeloma;
XX Alzheimer's disease; IGE; allergy; asthma; drug detoxification;
XX autoimmune; inflammatory disease; gene therapy; strep-tag.
XX
XX OS Synthetic.
XX
XX PN WO915563-A1.
XX
XX PD 01-APR-1999.
XX
XX PF 18-SEP-1998; 98WO-AU000783.
XX
XX PR 19-SEP-1997; 97AU-00009306.

```

```

XX
XX (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
XX PA Koentgen F, Suesse GM, Tarlinton DM, Treutlein HR;
XX
XX PI WPI; 1999-244394/20.
XX
XX DR Growth factor precursor cleaved by antigen-specific catalytic antibody.
XX
XX PT Example 3; Page 42; 101pp; English.
XX
XX PS
XX
XX The invention relates to a growth factor precursor that comprises B-cell
XX surface binding part, T cell surface binding part, antigen cleavable by a
XX catalytic antibody (Cab); and a peptide comprising heavy and light chains
XX of immunoglobulin. When the antigen is cleaved the B cell surface part
XX can interact with its target. The growth factor precursors are used to
XX select B cells that produce Ag-specific Cab, and to generate Cab from
XX such cells (by inducing mitogenesis, caused by the growth factor released
XX by specific cleavage). The Ag-specific Cab can be directed against, e.g.
XX tumour necrosis factor (treatment of rheumatoid arthritis or toxic shock
XX syndrome); viral docking receptors (treatment of human immune virus,
XX hepatitis and influenza infections); tumour-specific antigens; amyloid
XX plaque (treatment of Alzheimer's disease or myeloma) or IGE (treatment of
XX allergies such as asthma). Cab may also be used for drug detoxification,
XX to treat autoimmune or inflammatory diseases and to eliminate
XX environmental or industrial pollutants, such as plastics and petroleum.
XX Particularly the growth factor precursors are produced by delivering the
XX corresponding nucleic acid in a viral or other gene therapy vector. The
XX present sequence represents a strep-tag
XX
XX Sequence 9 AA;
XX
XX Query Match 88.6%; Score 39; DB 2; Length 9;
XX Best Local Similarity 85.7%; Pred. No. 1.8e+06;
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 XHPOFGG 8
XX :|||||
XX 3 RHPQFGG 9
XX
XX Db
XX
XX RESULT 6
XX AAB30795
XX ID AAB30795 standard; peptide; 9 AA.
XX
XX AC AAB30795;
XX
XX DT 02-APR-2001 (first entry)
XX
XX DE Amino acid sequence of a Strep epitope.
XX
XX KW SCHAG; self-coalesce; higher-order aggregate; amyloidogenic domain;
XX aggregation; fibril; phenotypic alteration; gene therapy;
XX disease resistance; plant pigmentation; prion disease.
XX
XX OS Synthetic.
XX
XX PN WO200075324-A2.
XX
XX PD 14-DEC-2000.
XX
XX PF 09-JUN-2000; 2000WO-US015876.
XX
XX PR 09-JUN-1999; 99US-0138833P.
XX
XX PA (ARCH-) ARCH DEV CORP.
XX
XX PI Lindquist S, Li L, Ma J, Liu J, Sontheimer N, Scheibel T;
XX
XX DR WPI; 2001-061723/07.
XX
XX PT New nucleic acid encoding chimeric proteins with self-assembly
XX properties, useful e.g. for diagnosis and treatment of prion diseases,

```


Query Match 88.6%; Score 39; DB 5; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
:|||||
DB 3 RHPQFG 9

RESULT 9
AA016086
ID AA016086 standard; peptide; 9 AA.

AC AA016086;
XX 27-FEB-2003 (first entry)
XX
XX Neurological/CNS disease treatment method-related peptide #18.

XX Vaccine; gene therapy; neurological disease; CNS disorder;
XX central nervous system disorder; olfactory system; Alzheimer's disease;
XX Creutzfeldt-Jakob disease; Huntington's chorea; Parkinson's disease;
XX viral infection of the brain; brain tumour; lysosomal storage disease;
XX multiple sclerosis.

OS Unidentified.
XX
XX WO200274243-A2.

XX 26-SEP-2002.

XX 15-MAR-2002; 2002WO-US008042.

XX 15-MAR-2001; 2001US-00808037.

XX (UYRA-) UNIT RAMOT APPLIED RES & IND DEV LTD.
XX (MCIN/) MCINNTS P.

XX Solomon B, Frenkel D;

XX WPI; 2003-040542/03.

XX Treating or diagnosing neurological diseases of the central nervous
XX system, e.g. Alzheimer's disease, comprises displaying a polypeptide or
XX diagnostic agent on viral display vehicle and introducing or detecting
XX the display vehicle.

XX Disclosure; Page 214; 214pp; English.

XX The invention comprises a method for treating a neurological disease or a
XX central nervous system (CNS) disorder. The method involves displaying a
XX therapeutic molecule capable of treating the neurological disease or CNS
XX disorder on a viral display vehicle. The viral display vehicle is then
XX introduced into the olfactory system of a subject to treat the disease or
XX disorder. The method of the invention is useful for preventing, treating
XX and diagnosing neurological diseases or CNS disorders, such as:
XX Alzheimer's disease; Creutzfeldt-Jakob disease; Huntington's chorea; viral
XX infections of the brain; brain tumours; lysosomal storage diseases;
XX Parkinson's disease; and multiple sclerosis. The present amino acid
XX sequence represents a peptide which was used in the invention

XX Sequence 9 AA;

Query Match 88.6%; Score 39; DB 6; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
:|||||
DB 3 RHPQFG 9

RESULT 10

ABP5547
ID ABP5547 standard; peptide; 9 AA.

XX ABP5547;

XX 19-FEB-2003 (first entry)

XX Streptag peptide SEQ ID NO:77.

XX Hepatitis C virus; HCV; envelope protein; hepatotropic; virucide;
XX antiinflammatory; gene therapy; vaccine.

XX Hepatitis C virus.
XX Synthetic.

XX WO200285932-A2.

XX 31-OCT-2002.

XX 24-APR-2002; 2002WO-BE000062.

XX 24-APR-2001; 2001EP-00870088.

XX 17-JUL-2001; 2001US-0305604P.

XX (INNO-) INNOGENETICS NV.

XX Sablon E, Van Broekhoven A, Bosman A, Depla E, Dechamps G;

XX WPI; 2003-093095/08.

XX New recombinant nucleic acids for expressing Hepatitis C virus (HCV)
XX envelope proteins in eukaryotic cells, comprising a sequence encoding a
XX protein having an avian lysozyme leader peptide joined to the HCV
XX envelope protein.

XX Claim 5; Page 303; 319pp; English.

XX The present invention describes a recombinant nucleic acid (I) comprising
XX a nucleotide sequence encoding a protein having an avian lysozyme leader
XX peptide, or its functional equivalent, joined to a Hepatitis C virus
XX (HCV) envelope protein or its part. Also described: (1) a vector
XX comprising the recombinant nucleic acid; (2) a host cell comprising the
XX recombinant nucleic acid or the vector; and (3) a method for producing
XX HCV envelope protein or its part in a host cell, comprising transforming
XX the host cell with the recombinant nucleic acid or with the vector, where
XX the host cell is capable of expressing the protein cited above. (I) has
XX hepatotropic, virucide and antiinflammatory activities, and can be used
XX in gene therapy and vaccines. The recombinant nucleic acid is useful for
XX efficient expression of Hepatitis C virus envelope proteins in eukaryotic
XX cells, such as yeast cells. The HCV envelope proteins may be used as a
XX vaccine, or for incorporation into an immunoassay for the detection of
XX anti-HCV antibodies, and/or genotyping of HCV, for prognosing or
XX monitoring of HCV disease, or as a therapeutic agent. The method is used
XX for producing HCV envelope protein or its part in a host cell. AB084197
XX to AB084253 and ABP5528 to ABP5568 represent sequences used in the
XX exemplification of the present invention

XX Sequence 9 AA;

Query Match 88.6%; Score 39; DB 6; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
:|||||
DB 3 RHPQFG 9

RESULT 11

AAE37229
ID AAE37229 standard; peptide; 9 AA.

XX AAE37229;

XX 07-AUG-2003 (first entry)
DT Strep-tag peptide.
XX
XX Gene expression; therapy; isolation.
XX
XX Synthetic.
XX
XX WO2003038049-A2.
XX
XX 08-MAY-2003.
XX
XX 29-OCT-2002; 2002WO-US034645.
XX
XX 29-OCT-2001; 2001US-0340689P.
XX
XX (RENO-) RENOVIS INC.
XX
XX Heintz N, Serafini TA, Shyjan AW;
XX
XX WPI; 2003-430512/40.
XX
XX Isolating cell-type specific mRNAs, useful in gene expression analysis or
PT quantification in a specific cell in a heterogeneous cell mixture, by
PT isolating mRNA bound to ribosomes or proteins that bind mRNA in a cell
PT type specific manner.
XX
XX
XX Disclosure; Page 8; 136pp; English.
XX
XX The invention relates to a method for isolating mRNA from a population of
CC cells. The method involves selectively isolating ribosomes or proteins
CC that bind mRNA in a cell type specific manner and then isolating the mRNA
CC bound to the ribosomes or proteins that bind mRNA. The method is useful
CC for facilitating the analysis and quantification of gene expression in a
CC selected cell type present within a heterogeneous cell mixture. The
CC method may also be used in diagnostics or therapies for human diseases.
CC The present sequence is Strep-tag peptide. This sequence is used to
CC illustrate the method of the invention
XX
XX Sequence 9 AA;
SQ

Query Match 88.6%; Score 39; DB 6; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 XHPQFGG 8
DB 3 RHPQFGG 9

RESULT 12
AAE32860
ID AAE32860 standard; peptide; 9 AA.
XX
XX AAE32860;
AC
XX
XX 24-MAR-2003 (first entry)
DT
XX
XX Streptag peptide.
DE
XX
XX Hepatitis C virus; HCV; envelope protein; virucide; immunostimulant;
KW immune response; T-cell; therapy; infection; pharmaceutical; vaccine.
XX
XX Synthetic.
OS
XX
XX WO200286101-A2.
XX
XX 31-OCT-2002.
PD
XX
XX 24-APR-2002; 2002WO-BE000064.
PF
XX
XX 24-APR-2001; 2001EP-00870088.
PR

PR 17-JUL-2001; 2001US-0305604P.
XX
XX (INNO-) INNOGENETICS NV.
XX
XX Depia E, Bosman A, Deschamps G, Sablon E, Suckow M, Samson I;
PI Verheyden G;
XX
XX WPI; 2003-103409/09.
XX
XX New Hepatitis C virus (HCV) envelope protein with N-glycosylation
PT site(s), useful as a vaccine for inducing a HCV-specific immune response
PT or HCV-specific antibodies, particularly for preventing or treating HCV
PT infection.
XX
XX Claim 10; Page 336; 355pp; English.
XX
XX The invention relates to Hepatitis C virus (HCV) envelope protein with N-
CC glycosylation site(s). The HCV envelope protein, or the pharmaceutical
CC composition comprising the envelope protein, is useful as a medicament or
CC a vaccine, particularly for inducing a HCV-specific immune response,
CC inducing HCV-specific antibodies or inducing a T-cell function in a
CC mammal. The protein is particularly useful for preventing, treating or
CC diagnosing HCV infection. It is also useful for detecting the presence of
CC anti-HCV antibodies in a sample. The present sequence is a peptide used
CC in the invention
XX
XX Sequence 9 AA;
SQ

Query Match 88.6%; Score 39; DB 6; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 XHPQFGG 8
DB 3 RHPQFGG 9

RESULT 13
AAE33270
ID AAE33270 standard; peptide; 9 AA.
XX
XX AAE33270;
AC
XX
XX 02-APR-2003 (first entry)
DT
XX
XX Strep epitope tag used to illustrate the method of the invention.
DE
XX
XX Membrane-spanning signal-transducing protein; MSST protein; epitope.
KW
XX
XX Synthetic.
OS
XX
XX WO200286507-A1.
XX
XX 31-OCT-2002.
PD
XX
XX 24-APR-2002; 2002WO-US013250.
PF
XX
XX 24-APR-2001; 2001US-0286250P.
PR 21-AUG-2001; 2001US-00935061.
XX
XX (STRD) UNIV LELAND STANFORD JUNIOR.
PA
XX
XX Koblikka BK, Ghanouni P, Lee TW;
PI
XX
XX WPI; 2003-103418/09.
DR
XX
XX Identifying an agent that modulates activity of a membrane-spanning,
PT signal-transducing (MSST) protein, by detecting a conformational change
PT in a MSST protein upon interaction with a ligand.
XX
XX Disclosure; Page 84; 104pp; English.
PS
XX
XX The present invention relates to methods and compositions for identifying

CC agents that modulate activity of a membrane-spanning, signal-transducing (MST) protein. The method involves detecting a conformational change in a MST protein upon interaction with a ligand. The method is useful for identifying agents that modulate (e.g. agonists or antagonists) activity of MST protein. The present sequence is a strep epitope tag used to illustrate the method of the invention

XX
SQ Sequence 9 AA;

Query Match 88.6%; Score 39; DB 6; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPORFG 8
:|||||
Db 3 RHPORFG 9

RESULT 14
ABG72479 standard; peptide; 9 AA.

XX
AC ABG72479;
XX
DT 18-FEB-2003 (first entry)
XX
DE Strep tag for use as a G protein epitope tag.
XX
XX G protein; alpha sub-unit; G protein coupled receptor; GPCR;
KW G protein activation; G protein coupled receptor activation;
KM G protein coupled receptor identification; strep tag.
XX
OS Synthetic.
XX
PN US6448377-B1.
XX
PD 10-SEP-2002.
XX
PF 27-SEP-2000; 2000US-00672239.
XX
PR 27-SEP-2000; 2000US-00672239.
XX
PA (STRD) UNIV LELAND STANFORD JUNIOR.
XX
PI Koblika B, Lee TW;
XX
DR WPI; 2003-110149/10.
XX
PT Determining effects of candidate agent on activation of a G protein coupled receptor (GPCR) for evaluating new agonists and/or inverse agonists for GPCR by contacting a candidate agent with a modified G protein alpha subunit and a GPCR.
XX
PS Disclosure; Col 9; 38pp; English.
XX
CC The invention describes a method of determining the effects of a candidate agent on activation of a G protein coupled receptor (GPCR). The method comprises contacting a candidate agent with a modified G protein alpha subunit and a GPCR, and detecting a level of G protein activation in response to the contacting, where the level of activation is indicative of the effects of the agent on the activity of GPCR. The method is useful for determining the effects of a candidate agent on activation of a G protein coupled receptor, evaluating new agonists, and/or inverse agonists for GPCRs, identifying ligands for GPCRs, and developing a strategy for identifying GPCRs involved in different biological processes, including diseases. The invention provides rapid and more sensitive bioassays for evaluating new agonists, agonists and/or inverse agonists for GPCRs. The method can be performed using membranes, which increases both the ease of performing the assay and its efficacy, and also allows high throughput screening of GPCR activity. Furthermore, this method directly measures GPCR activity, and thus is less labour-intensive than the conventional methods. This sequence represents the Strep tag, an example of an epitope tag that can be used to tether the

CC stimulatory G protein alpha sub-unit to a membrane

XX
SQ Sequence 9 AA;

Query Match 88.6%; Score 39; DB 6; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPORFG 8
:|||||
Db 3 RHPORFG 9

RESULT 15
ADB84587 standard; peptide; 9 AA.

XX
AC ADB84587;
XX
DT 04-DEC-2003 (first entry)
XX
DE Streptavidin conserved peptide #1.
XX
KW cell-free transcription system; cell-free translation system;
KM protein synthesis; matrix; streptavidin.
XX
OS Escherichia coli.
XX
PN DE10137792-A1.
XX
PD 27-FEB-2003.
XX
PF 06-AUG-2001; 2001DE-01037792.
XX
PR 06-AUG-2001; 2001DE-01037792.
XX
PA (ERDM/) ERDMANN V.
XX
PI Erdmann VA, Lamla T, Stiege W;
XX
DR WPI; 2003-143999/33.
XX
PT Expressing genes in cell-free system, useful for preparation of proteins, comprises that the protein formed is removed from solution by binding to a matrix.
XX
PS Claim 13; Col 8; 8pp; German.
XX
CC This invention describes a novel method of expressing genes in a cell-free transcription and translation system which comprises using a reaction solution containing all necessary components of the transcription/translation system, amino acids, nucleotides and metabolites that supply energy and that are needed for synthesis. The proteins formed are immobilised on a matrix. The method allows simple recovery of proteins without a separate isolation step and the amount of proteins produced can be determined before a reaction is complete.
CC Continuous removal of proteins prevents it interfering with the expression process, making possible synthesis of proteins that interact adversely with the process, so normally produced only in very low yields.
CC The use of a matrix concentrates the proteins produced and is applicable to proteins of any size. This sequence represents a highly conserved peptide from E. coli streptavidin which is used to illustrate the method of the invention.
XX
SQ Sequence 9 AA;

Query Match 88.6%; Score 39; DB 7; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPORFG 8
:|||||
Db 3 RHPORFG 9

Wed Mar 2 14:27:08 2005

seq4.rag

Page 8

Search completed: March 2, 2005, 13:02:44
JOB time : 36.0325 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 24.878 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ4

Perfect score: 44

Sequence: 1 kxhpfgg 8

Scoring table: BLOSUM62X

Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	88.6	8	14	US-10-026-578B-1
2	39	88.6	9	9	US-09-808-037-32
3	39	88.6	9	9	US-09-803-067-2
4	39	88.6	9	10	US-09-935-061-3
5	39	88.6	9	14	US-10-128-590-77
6	39	88.6	9	14	US-10-345-618-14
7	39	88.6	9	14	US-10-128-587A-77
8	39	88.6	9	15	US-10-384-788-32
9	39	88.6	9	15	US-10-449-831A-166
10	39	88.6	9	15	US-10-425-000-75
11	39	88.6	9	15	US-10-424-999-24
12	39	88.6	9	15	US-10-272-196-30
13	39	88.6	9	16	US-10-612-410-18

14	39	88.6	9	16	US-10-692-071-3	Sequence 3, Appli
15	39	88.6	9	16	US-10-397-438A-6	Sequence 6, Appli
16	39	88.6	9	16	US-10-753-309-5	Sequence 5, Appli
17	39	88.6	9	17	US-10-494-248-5	Sequence 5, Appli
18	39	88.6	10	9	US-09-801-968-18	Sequence 18, Appli
19	39	88.6	10	10	US-09-802-154-18	Sequence 18, Appli
20	39	88.6	10	13	US-10-060-765-13	Sequence 13, Appli
21	39	88.6	10	14	US-10-263-230A-12	Sequence 12, Appli
22	39	88.6	10	16	US-10-818-140-13	Sequence 13, Appli
23	39	88.6	10	17	US-10-771-173-13	Sequence 13, Appli
24	39	88.6	12	9	US-09-904-599A-8	Sequence 8, Appli
25	39	88.6	18	16	US-10-397-438A-9	Sequence 9, Appli
26	39	88.6	20	15	US-10-342-805-1	Sequence 1, Appli
27	39	88.6	23	16	US-10-344-607-2	Sequence 2, Appli
28	39	88.6	42	16	US-10-344-607-16	Sequence 16, Appli
29	39	88.6	42	16	US-10-344-607-20	Sequence 20, Appli
30	39	88.6	46	15	US-10-422-262-12	Sequence 12, Appli
31	39	88.6	46	15	US-10-448-609-8	Sequence 8, Appli
32	39	88.6	47	15	US-10-416-230-57	Sequence 57, Appli
33	39	88.6	132	15	US-10-449-831A-212	Sequence 212, App
34	39	88.6	155	14	US-10-345-618-11	Sequence 11, Appli
35	39	88.6	178	14	US-10-345-618-13	Sequence 13, Appli
36	39	88.6	194	16	US-10-416-708A-10	Sequence 10, Appli
37	39	88.6	194	16	US-10-416-708A-64	Sequence 64, Appli
38	39	88.6	198	14	US-10-345-618-8	Sequence 8, Appli
39	39	88.6	210	9	US-09-272-809-24	Sequence 24, Appli
40	39	88.6	212	15	US-10-449-831A-198	Sequence 198, App
41	39	88.6	234	15	US-10-449-831A-192	Sequence 192, App
42	39	88.6	329	15	US-10-449-831A-204	Sequence 204, App
43	39	88.6	334	13	US-10-017-736-4	Sequence 4, Appli
44	39	88.6	334	15	US-10-650-585-4	Sequence 4, Appli
45	39	88.6	342	14	US-10-345-618-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-10-026-578B-1
; Sequence 1, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent version 3.1
; SEQ ID NO 1
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-1

Query Match 88.6%; Score 39; DB 14; Length 8;
Best local similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 KXHPFGG 8
:|||||
DB 2 RHHPFGG 8

```
RESULT 2
US-09-808-037-32
; Sequence 32, Application US/0908037
; Patent No. US20020052311A1
; GENERAL INFORMATION:
; APPLICANT: SOLOMON, Beke
; APPLICANT: HANAN, Eliat
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE TREATMENT AND/OR DIAGNOSIS OF
; FILE REFERENCE: SOLOMON-2D
; CURRENT APPLICATION NUMBER: US/09/808,037
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: 09/629,971
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 09/473,653
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: US 60/152,417
; PRIOR FILING DATE: 1999-09-03
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 32
; LENGTH: 9
; TYPE: PRT
; ORGANISM: synthetic construct
US-09-808-037-32

Query Match      88.6%; Score 39; DB 9; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
DB      3 RHPQFG 9

RESULT 3
US-09-983-067-2
; Sequence 2, Application US/09983067
; Patent No. US20020123101A1
; GENERAL INFORMATION:
; APPLICANT: INOUE, Akio
; APPLICANT: SHIMIZU, Yoshihiro
; APPLICANT: UEDA, Takuya
; TITLE OF INVENTION: Process for Producing Peptides By Using In Vitro
; FILE REFERENCE: 1752-0151P
; CURRENT APPLICATION NUMBER: US/09/983,067
; CURRENT FILING DATE: 2001-10-23
; PRIOR APPLICATION NUMBER: JP 294795/2001
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: JP 227094/2001
; PRIOR FILING DATE: 2001-07-27
; PRIOR APPLICATION NUMBER: JP 6910/2001
; PRIOR FILING DATE: 2001-01-15
; PRIOR APPLICATION NUMBER: JP 401417/2000
; PRIOR FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Strept-tag binding to streptavidin
; LOCATION: (1)..(9)
; OTHER INFORMATION: Schmidt & Skerra, 1993, "The random peptide library-assisted
; OTHER INFORMATION: engineering of a C-terminal affinity peptide, useful for the
; OTHER INFORMATION: detection and purification of a functional IgFv fragment",
US-09-983-067-2
```

```
Query Match      88.6%; Score 39; DB 9; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
DB      3 RHPQFG 9

RESULT 4
US-09-935-061-3
; Sequence 3, Application US/09935061
; Publication No. US20030129649A1
; GENERAL INFORMATION:
; APPLICANT: Koblika, Brian M.
; APPLICANT: Chanouni, Pejman
; APPLICANT: Lee, Tae Weon
; TITLE OF INVENTION: Conformational assays to detect binding
; FILE REFERENCE: STAN213
; CURRENT APPLICATION NUMBER: US/09/935,061
; CURRENT FILING DATE: 2001-08-21
; PRIOR APPLICATION NUMBER: 60/286,250
; PRIOR FILING DATE: 2001-04-24
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: epitope tag peptide
US-09-935-061-3

Query Match      88.6%; Score 39; DB 10; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
DB      3 RHPQFG 9
```

```
RESULT 5
US-10-128-590-77
; Sequence 77, Application US/10128590
; Publication No. US20030108561A1
; GENERAL INFORMATION:
; APPLICANT: Innogenetics N.V.
; TITLE OF INVENTION: Core-glycosylated HCV envelope proteins
; FILE REFERENCE: 135 PCT
; CURRENT APPLICATION NUMBER: US/10/128,590
; CURRENT FILING DATE: 2002-07-22
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77
; LENGTH: 9
; TYPE: PRT
; ORGANISM: streptag
US-10-128-590-77

Query Match      88.6%; Score 39; DB 14; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
DB      3 RHPQFG 9

RESULT 6
US-10-345-618-14
; Sequence 14, Application US/10345618
```



```
/ Publication No. US2003014848A1
/ GENERAL INFORMATION:
/ APPLICANT: Koentgen, Frank
/ APPLICANT: Suesse, Gabriele M.
/ APPLICANT: Tarlinton, David M.
/ APPLICANT: Treutlein, Herbert R.
/ TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF PRODUCING SAME
/ FILE REFERENCE: 13474
/ CURRENT APPLICATION NUMBER: US/10/345,618
/ CURRENT FILING DATE: 2003-01-16
/ PRIOR APPLICATION NUMBER: US/09/509,031
/ PRIOR FILING DATE: 2000-06-09
/ NUMBER OF SEQ ID NOS: 16
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 14
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence:Strep-tag
US-10-345-618-14

Query Match      88.6%; Score 39; DB 14; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 7
US-10-128-587A-77
/ Sequence 77, Application US/10128587A
/ Publication No. US20030152940A1
/ GENERAL INFORMATION:
/ APPLICANT: Indogene N.V.
/ TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope
/ FILE REFERENCE: 134 PCT
/ CURRENT APPLICATION NUMBER: US/10/128,587A
/ CURRENT FILING DATE: 2002-04-24
/ NUMBER OF SEQ ID NOS: 98
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 77
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: streptag
US-10-128-587A-77

Query Match      88.6%; Score 39; DB 14; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 8
US-10-384-788-32
/ Sequence 32, Application US/10384788
/ Publication No. US20040013647A1
/ GENERAL INFORMATION:
/ APPLICANT: SOLOMON, Beke
/ APPLICANT: FRENKEL, Dan
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING A PLAQUE-FORMING DISEASE
/ FILE REFERENCE: SOLOMON-2D-2
/ CURRENT APPLICATION NUMBER: US/10/384,788
/ CURRENT FILING DATE: 2003-03-11
/ PRIOR APPLICATION NUMBER: 60/371,735
```

```
/ PRIOR FILING DATE: 2002-04-12
/ PRIOR APPLICATION NUMBER: 09/808,037
/ PRIOR FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: 09/830,954
/ PRIOR FILING DATE: 2001-06-22
/ PRIOR APPLICATION NUMBER: 10/162,889
/ PRIOR FILING DATE: 2002-06-06
/ PRIOR APPLICATION NUMBER: 09/473,653
/ PRIOR FILING DATE: 1999-12-29
/ PRIOR APPLICATION NUMBER: 09/629,971
/ PRIOR FILING DATE: 2000-07-31
/ PRIOR APPLICATION NUMBER: 60/152,417
/ PRIOR FILING DATE: 1999-09-03
/ PRIOR APPLICATION NUMBER: PCT/IL00/00518
/ PRIOR FILING DATE: 2000-08-31
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 32
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Synthetic construct
US-10-384-788-32

Query Match      88.6%; Score 39; DB 15; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 9
US-10-449-831A-166
/ Sequence 166, Application US/10449831A
/ Publication No. US20040029179A1
/ GENERAL INFORMATION:
/ APPLICANT: Koentgen, Frank
/ TITLE OF INVENTION: Higher molecular weight entities and uses therefor
/ FILE REFERENCE: 2385978
/ CURRENT APPLICATION NUMBER: US/10/449,831A
/ CURRENT FILING DATE: 2003-05-30
/ PRIOR APPLICATION NUMBER: USSN 60/384878
/ PRIOR FILING DATE: 2002-05-31
/ NUMBER OF SEQ ID NOS: 237
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 166
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Strep tag
US-10-449-831A-166

Query Match      88.6%; Score 39; DB 15; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 10
US-10-425-000-75
/ Sequence 75, Application US/10425000
/ Publication No. US20040052777A1
/ GENERAL INFORMATION:
/ APPLICANT: Nesbitt, Mark
/ APPLICANT: Cameron, Beatrice
/ APPLICANT: Blanche, Francis
/ TITLE OF INVENTION: Kringle Polypeptides and Methods for Using Them to Inhibit
/ TITLE OF INVENTION: Angiogenesis
```

FILE REFERENCE: ST01027-B
CURRENT APPLICATION NUMBER: US/10/425,000
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: 10/233,675
PRIOR FILING DATE: 2002-09-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: PatentIn version 3.2
SEQ ID NO 75
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Purification tag
US-10-425-000-75

Query Match 88.6%; Score 39; DB 15; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFGG 8
:|||||
Db 3 RHPQFGG 9

RESULT 11
US-10-424-999-24
Sequence 24, Application US/10424999
Publication No. US20040052810A1
GENERAL INFORMATION:
APPLICANT: Nesbitt, Mark
APPLICANT: Cameron, Beatrice
APPLICANT: Blanche, Francis
TITLE OF INVENTION: Abrogen Polypeptides, Nucleic Acids Encoding Them and Methods for
FILE REFERENCE: ST01027-A
CURRENT APPLICATION NUMBER: US/10/424,999
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: 10/233,675
PRIOR FILING DATE: 2002-09-04
NUMBER OF SEQ ID NOS: 70
SOFTWARE: PatentIn version 3.2
SEQ ID NO 24
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Purification tag
US-10-424-999-24

Query Match 88.6%; Score 39; DB 15; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFGG 8
:|||||
Db 3 RHPQFGG 9

RESULT 12
US-10-272-196-30
Sequence 30, Application US/10272196
Publication No. US20040072746A1
GENERAL INFORMATION:
APPLICANT: Murray, Clare Margaret
APPLICANT: Hutchinson, Raymond
APPLICANT: Bartick, John Raymond
APPLICANT: Sullivan, Michael
APPLICANT: Donald, David Keith
APPLICANT: Jackson, Andrew Paul
APPLICANT: Cook, Ian David
TITLE OF INVENTION: INHIBITORS OF MONOCARBOXYLATE TRANSPORT
FILE REFERENCE: 06275-285001

CURRENT APPLICATION NUMBER: US/10/272,196
CURRENT FILING DATE: 2002-10-16
PRIOR APPLICATION NUMBER: US 60/329,318
PRIOR FILING DATE: 2001-10-16
NUMBER OF SEQ ID NOS: 44
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 30
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
US-10-272-196-30

Query Match 88.6%; Score 39; DB 15; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFGG 8
:|||||
Db 3 RHPQFGG 9

RESULT 13
US-10-612-410-18
Sequence 18, Application US/10612410
Publication No. US20040132133A1
GENERAL INFORMATION:
APPLICANT: Bennett, Robert P.
TITLE OF INVENTION: Methods and Compositions for the Production, Identification and
FILE REFERENCE: 0942,5510003
CURRENT APPLICATION NUMBER: US/10/612,410
PRIOR APPLICATION NUMBER: 2003-07-03
PRIOR FILING DATE: 2002-07-08
PRIOR APPLICATION NUMBER: 60/393,756
PRIOR FILING DATE: 2002-07-19
PRIOR APPLICATION NUMBER: 60/396,627
PRIOR FILING DATE: 2002-07-19
PRIOR APPLICATION NUMBER: 60/417,172
PRIOR FILING DATE: 2002-10-10
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn version 3.2
SEQ ID NO 18
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Strep epitope
US-10-612-410-18

Query Match 88.6%; Score 39; DB 16; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFGG 8
:|||||
Db 3 RHPQFGG 9

RESULT 14
US-10-692-071-3
Sequence 3, Application US/10692071
Publication No. US20040157268A1
GENERAL INFORMATION:
APPLICANT: Koblika, Brian K.
APPLICANT: Chanount, Pejman
APPLICANT: Lee, Tae Meon
TITLE OF INVENTION: Conformational assays to detect binding
FILE REFERENCE: STAN-213CIP
CURRENT APPLICATION NUMBER: US/10/692,071
CURRENT FILING DATE: 2003-10-22
PRIOR APPLICATION NUMBER: PCT/US02/13250
PRIOR FILING DATE: 2002-04-24
PRIOR APPLICATION NUMBER: 09/935,061

PRIOR FILING DATE: 2001-08-21
PRIOR APPLICATION NUMBER: 60/286,250
PRIOR FILING DATE: 2001-04-24
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: epitope tag peptide
US-10-692-071-3

Query Match 88.6%; Score 39; DB 16; length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
:|||||
Db 3 RHPQFG 9

RESULT 15
US-10-397-438A-6
Sequence 6, Application US/10397438A
Publication No. US20040191869A1
GENERAL INFORMATION:
APPLICANT: Pharmacia & Upjohn AB
TITLE OF INVENTION: Fusion vectors
FILE REFERENCE: 00126
CURRENT APPLICATION NUMBER: US/10/397,438A
CURRENT FILING DATE: 2002-11-05
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Strep-tag
US-10-397-438A-6

Query Match 88.6%; Score 39; DB 16; length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
:|||||
Db 3 RHPQFG 9

Search completed: March 2, 2005, 14:18:50
Job time : 25.874 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 9.10569 Seconds

(without alignment)
65.585 Million cell updates/sec

Title: SEQ4

Perfect score: 44

Sequence: 1 kxhpqfsg 8

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: Issued Patents AA:
2: /cgn2_6/prodata/1/aa/5A_COMB.pep.*
3: /cgn2_6/prodata/1/aa/6A_COMB.pep.*
4: /cgn2_6/prodata/1/aa/6B_COMB.pep.*
5: /cgn2_6/prodata/1/aa/PCTUS_COMB.pep.*
6: /cgn2_6/prodata/1/aa/backfile1.pep.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	88.6	9	1	US-08-660-626-3
2	39	88.6	9	3	US-08-828-741B-14
3	39	88.6	9	3	US-08-948-097-1
4	39	88.6	9	3	US-09-031-168-3
5	39	88.6	9	3	US-09-160-567-14
6	39	88.6	9	4	US-09-672-239-3
7	39	88.6	9	4	US-09-710-299-14
8	39	88.6	9	4	US-09-509-031-14
9	39	88.6	9	4	US-09-669-516C-3
10	39	88.6	10	1	US-08-294-386C-11
11	39	88.6	10	2	US-08-737-316A-5
12	39	88.6	10	3	US-08-897-020-6
13	39	88.6	10	3	US-08-895-707-11
14	39	88.6	10	3	US-09-350-823-6
15	39	88.6	10	4	US-09-715-805-13
16	39	88.6	10	4	US-09-692-845-9
17	39	88.6	10	4	US-09-640-041-7
18	39	88.6	10	5	PCT-US95-10224-11
19	39	88.6	12	4	US-08-218-369-8
20	39	88.6	12	4	US-09-304-599A-8
21	39	88.6	12	5	PCT-US95-03742-8
22	39	88.6	15	1	US-08-664-449-33
23	39	88.6	155	3	US-08-828-741B-11
24	39	88.6	155	3	US-09-160-567-11
25	39	88.6	155	4	US-09-710-299-11
26	39	88.6	155	4	US-09-509-031-11
27	39	88.6	178	3	US-08-828-741B-13

28	39	88.6	178	3	US-09-160-567-13	Sequence 13, Appl
29	39	88.6	178	4	US-09-710-299-13	Sequence 13, Appl
30	39	88.6	178	4	US-09-509-031-13	Sequence 13, Appl
31	39	88.6	197	3	US-08-897-020-7	Sequence 7, Appl
32	39	88.6	197	3	US-09-350-823-7	Sequence 7, Appl
33	39	88.6	198	3	US-08-828-741B-8	Sequence 8, Appl
34	39	88.6	198	3	US-09-160-567-8	Sequence 8, Appl
35	39	88.6	198	4	US-09-710-299-8	Sequence 8, Appl
36	39	88.6	198	4	US-09-509-031-8	Sequence 8, Appl
37	39	88.6	334	4	US-10-017-736C-4	Sequence 4, Appl
38	39	88.6	342	3	US-08-828-741B-6	Sequence 6, Appl
39	39	88.6	342	3	US-09-160-567-6	Sequence 6, Appl
40	39	88.6	342	4	US-09-710-299-6	Sequence 6, Appl
41	39	88.6	342	4	US-09-509-031-6	Sequence 6, Appl
42	39	88.6	386	3	US-08-895-707-7	Sequence 7, Appl
43	39	88.6	409	4	US-10-017-736C-2	Sequence 2, Appl
44	39	88.6	487	4	US-09-270-767-43055	Sequence 43055, A
45	39	88.6	495	3	US-08-828-741B-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-660-626-3
; Sequence 3, Application US/08660626
; Patent No. 5789655
; GENERAL INFORMATION:
; APPLICANT: Stanley B. Prusiner
; APPLICANT: Glenn C. Telling
; APPLICANT: Fred E. Cohen
; APPLICANT: Michael R. Scott
; TITLE OF INVENTION: TRANSGENIC ANIMALS EXPRESSING
; TITLE OF INVENTION: EPTOPE-TAGGED PROTEINS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: California
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Asclii
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/660,626
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Valeta Gregg
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 07532/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 322-5070
; TELEFAX: (415) 854-0875
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Peptide
; US-08-660-626-3

Query Match 88.6%; Score 39; DB 1; Length 9;
Best Local Similarity 85.7%; Pred.No. 4;le+05;
Matches 6; Conservative 1; Mismatches 0; Gaps 0;
QY 2 XHPQFG 8
:|||||
Db 3 RHPQFG 9

RESULT 2
US-08-828-741B-14
; Sequence 14, Application US/08828741B
; Patent No. 6043069
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Suesse, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/828,741B
; FILING DATE: 26-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: DIGIGLIO, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-828-741B-14

Query Match 88.6%; Score 39; DB 3; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
 :|||||
Db 3 RHPQFG 9

RESULT 3
US-08-948-097-1
; Sequence 1, Application US/08948097C
; Patent No. 6103493
; GENERAL INFORMATION:
; APPLICANT: Skerra, Arne
; APPLICANT: Voss, Selma
; TITLE OF INVENTION: Streptavidin Muteins
; FILE REFERENCE: HUBR 1119
; CURRENT APPLICATION NUMBER: US/08/948,097C
; CURRENT FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: DE 196 41 876.3
; EARLIER FILING DATE: 1996-10-10
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 1
; LENGTH: 9
; TYPE: PRT

ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: BINDING
; OTHER INFORMATION: Binding ligand for streptavidin
US-08-948-097-1

Query Match 88.6%; Score 39; DB 3; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
 :|||||
Db 3 RHPQFG 9

RESULT 4
US-09-031-168-3
; Sequence 3, Application US/09031168
; Patent No. 6150583
; GENERAL INFORMATION:
; APPLICANT: Stanley B. Prusiner
; APPLICANT: Glenn C. Telling
; APPLICANT: Fred E. Cohen
; APPLICANT: Michael R. Scott
; TITLE OF INVENTION: TRANSGENIC ANIMALS EXPRESSING
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: California
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Asciti
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031,168
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/660,626
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Valeta Gregg
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 07532/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 322-5070
; TELEFAX: (415) 854-0875
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-031-168-3

Query Match 88.6%; Score 39; DB 3; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
 :|||||
Db 3 RHPQFG 9

RESULT 5
US-09-160-567-14
; Sequence 14, Application US/09160567

Patent No. 6326179
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Sues, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/160,567
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,741
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: DIGIGLO, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-160-567-14

Query Match 88.6%; Score 39; DB 3; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
; :|||||
; Db 3 RHPQFG 9

RESULT 6
US-09-672-239-3
; Sequence 3, Application US/09672239
; Patent No. 6448377
; GENERAL INFORMATION:
; APPLICANT: Lee, Tae Weon
; APPLICANT: Koblika, Brian
; TITLE OF INVENTION: MODIFIED G PROTEIN SUBUNITS
; FILE REFERENCE: STAN-204
; CURRENT APPLICATION NUMBER: US/09/672,239
; CURRENT FILING DATE: 2000-09-27
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-672-239-3

Query Match 88.6%; Score 39; DB 4; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
; :|||||
; Db 3 RHPQFG 9

RESULT 7
US-09-710-299-14
; Sequence 14, Application US/09710299
; Patent No. 6521741
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank
; APPLICANT: Sues, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Treutlein, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
; TITLE OF INVENTION: PRODUCING SAME
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
; STREET: 400 Garden City Plaza
; CITY: Garden City
; STATE: New York
; COUNTRY: United States of America
; ZIP: 11530
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/710,299
; FILING DATE: 09-NO. 6521741-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,741
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: DIGIGLO, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 10591
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-710-299-14

Query Match 88.6%; Score 39; DB 4; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
; :|||||
; Db 3 RHPQFG 9

RESULT 8
US-09-509-031-14
; Sequence 14, Application US/09509031
; Patent No. 6590080
; GENERAL INFORMATION:
; APPLICANT: Koentgen, Frank

APPLICANT: Sues, Gabriele M.
APPLICANT: Trenton, David M.
APPLICANT: Treutlein, Herbert R.
TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF PRODUCING SAME
FILE REFERENCE: 13474
CURRENT APPLICATION NUMBER: US/09/509,031
CURRENT FILING DATE: 2000-06-09
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 14
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Strep-tag
US-09-509-031-14

Query Match 88.6%; Score 39; DB 4; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 9
US-09-669-516C-3
Sequence 3, Application US/09669516C
Patent No. 6602672
GENERAL INFORMATION:
APPLICANT: Prubiner, Stanley B.
APPLICANT: Telling, Glenn C.
APPLICANT: Cohen, Fred E.
APPLICANT: Schitt, Michael R.
TITLE OF INVENTION: RECOMBINANT CONSTRUCT ENCODING EPTOPE
TITLE OF INVENTION: TAGGED PRP PROTEIN
FILE REFERENCE: UCAL-045CON
CURRENT APPLICATION NUMBER: US/09/669,516C
CURRENT FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: 09/031,168
PRIOR FILING DATE: 1998-02-26
PRIOR APPLICATION NUMBER: 08/660,626
PRIOR FILING DATE: 1996-06-06
PRIOR APPLICATION NUMBER: 08/521,992
PRIOR FILING DATE: 1995-08-31
PRIOR APPLICATION NUMBER: 08/509,261
PRIOR FILING DATE: 1995-07-31
PRIOR APPLICATION NUMBER: 08/242,188
PRIOR FILING DATE: 1994-05-13
NUMBER OF SEQ ID NOS: 15
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Strep peptide
US-09-669-516C-3

Query Match 88.6%; Score 39; DB 4; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 10
US-08-294-386C-11
Sequence 11, Application US/08294386C
Patent No. 5646030

GENERAL INFORMATION:
APPLICANT: Ray, Bryan L.
APPLICANT: Lin, Edmund C.C.
APPLICANT: Crea, Roberto
TITLE OF INVENTION: Method of Isolating Mutant Cells
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Lappin & Kusner
STREET: 200 State Street
CITY: Boston
STATE: MA
COUNTRY: U.S.A.
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/294,386C
FILING DATE: August 23, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kerner, Ann-Louise
REGISTRATION NUMBER: 33,523
REFERENCE/DOCKET NUMBER: STZZ-010CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/330-1300
TELEFAX: 617/330-1311
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 10 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
US-08-294-386C-11

Query Match 88.6%; Score 39; DB 1; Length 10;
Best Local Similarity 85.7%; Pred. No. 0.46;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 4 RHPQFG 10

RESULT 11
US-08-737-316A-5
Sequence 5, Application US/08737316A
Patent No. 5849576
GENERAL INFORMATION:
APPLICANT: SKERRA, Arne
APPLICANT: WARDENBERG, Christina
TITLE OF INVENTION: USE OF THE TETRACYCLINE PROMOTER FOR THE
TITLE OF INVENTION: STRINGENTLY REGULATED PRODUCTION OF RECOMBINANT PROTEINS IN PR
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSER: Nikaido, Marmelstein, Murray & Oram LLP
STREET: 655 Fifteenth Street N.W. Suite 330
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-5701
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/737,316A
FILING DATE: 12-NOV-1996


```
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/EP95/01862
/ FILING DATE: 17-MAY-1995
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: DE P 44 17 598.1
/ FILING DATE: 19-MAY-1994
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Kitter, Monica C.
/ REGISTRATION NUMBER: 36,105
/ REFERENCE/DOCKET NUMBER: P564-6012
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 638-5000
/ TELEFAX: (202) 638-4810
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 10 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
US-08-737-316A-5

Query Match      88.6%; Score 39; DB 2; Length 10;
Best Local Similarity 85.7%; Pred. No. 0.46;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPORFG 8
       :|||||
Db      4 RHPORFG 10

RESULT 12
US-08-897-020-6
/ Sequence 6, Application US/08897020
/ Patent No. 6028176
/ GENERAL INFORMATION:
/ APPLICANT: Shanfelt, Armen; Greve, Jeffrey; Rocznialk, Steven
/ TITLE OF INVENTION: High-affinity Interleukin-4 Mutains
/ NUMBER OF SEQUENCES: 9
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Bayer Corporation, Pharmaceutical Division
/ STREET: 400 Morgan Lane
/ CITY: West Haven
/ STATE: CT
/ COUNTRY: United States of America
/ ZIP: 06516-4175
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS V. 6.30
/ SOFTWARE: Word for Windows 6.0
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 10
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ DESCRIPTION: tag for streptavidin
/ HYPOTHETICAL: no
```

```
/ ANTI-SENSE: no
US-08-897-020-6

Query Match      88.6%; Score 39; DB 3; Length 10;
Best Local Similarity 85.7%; Pred. No. 0.46;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPORFG 8
       :|||||
Db      4 RHPORFG 10

RESULT 13
US-08-895-707-11
/ Sequence 11, Application US/08895707
/ Patent No. 6077700
/ GENERAL INFORMATION:
/ APPLICANT: (Pharmacia & Upjohn, Co.)
/ APPLICANT: alternatively, for U.S. filing:
/ APPLICANT: Hollingsworth, Robert A.
/ APPLICANT: Sharma, Satish K.
/ APPLICANT: Rank, Kenneth B.
/ APPLICANT: Evans, David B.
/ TITLE OF INVENTION: Special Constructs and Complexes of
/ TITLE OF INVENTION: Cyclin E
/ NUMBER OF SEQUENCES: 21
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Pharmacia & Upjohn Company
/ STREET: 301 Henrietta Street
/ CITY: Kalamazoo
/ STATE: MI
/ COUNTRY: USA
/ ZIP: 49001
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Rel. #1.0, Ver. #1.25/Wordperfect 5.2+
/ INFORMATION FOR SEQ ID NO: 11:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 10 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ HYPOTHETICAL: NO
/ ANTI-SENSE: NO
/ FRAGMENT TYPE: N-terminal
US-08-895-707-11

Query Match      88.6%; Score 39; DB 3; Length 10;
Best Local Similarity 85.7%; Pred. No. 0.46;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPORFG 8
       :|||||
Db      4 RHPORFG 10

RESULT 14
US-09-350-823-6
/ Sequence 6, Application US/09350823
```

Patent No. 6313272
GENERAL INFORMATION:
APPLICANT: Shanfelt, Armen; Greve, Jeffrey; Rocznik, Steven
TITLE OF INVENTION: High-affinity Interleukin-4 Mutains
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSER: Bayer Corporation, Pharmaceutical Division
STREET: 400 Morgan Lane
CITY: West Haven
STATE: CT
COUNTRY: United States of America
ZIP: 06516-4175
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS V. 6.30
SOFTWARE: Word for Windows 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/350,823
FILING DATE: 09-Jul-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/897,020
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Huw R. Jones
REGISTRATION NUMBER: 33, 916
REFERENCE/DOCKET NUMBER: WH5020
TELEPHONE: (203) 812-2317
TELEFAX: (203) 812-5492
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 10
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
DESCRIPTION: tag for streptavidin
HYPOTHETICAL: no
ANTI-SENSE: no
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-350-823-6
Query Match 88.6%; Score 39; DB 3; Length 10;
Best Local Similarity 85.7%; Pred. No. 0.46;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 XHPQFG 8
Db 4 RHPQFG 10
RESULT 15
US-09-715-805-13
Sequence 13, Application US/09715805
Patent No. 6716626
GENERAL INFORMATION:
APPLICANT: Itoh, No. 6716626uyuki
APPLICANT: Kavanaugh, W. Michael
TITLE OF INVENTION: HUMAN FGF-21 GENE AND GENE EXPRESSION
TITLE OF INVENTION: PRODUCTS
FILE REFERENCE: PP-16758.001/201130.408
CURRENT APPLICATION NUMBER: US/09/715,805
CURRENT FILING DATE: 2000-11-16
NUMBER OF SEQ ID NOS: 17
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 10
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Residues which bind to paramagnetic streptavidin

OTHER INFORMATION: beads (used for purification).
US-09-715-805-13
Query Match 88.6%; Score 39; DB 4; Length 10;
Best Local Similarity 85.7%; Pred. No. 0.46;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 XHPQFG 8
Db 4 RHPQFG 10
Search completed: March 2, 2005, 12:25:33
Job time: 9.10569 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 6.11382 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ5

Perfect score: 42

Sequence: 1 kxhpgfek 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 183416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Pir1:*
2: Pir2:*
3: Pir3:*
4: Pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	88.1	189	2 T36642	hypothetical prote
2	37	88.1	627	2 T00484	hypothetical prote
3	36	85.7	101	2 T10856	carboxypeptidase C
4	35	83.3	300	2 G75436	conserved hypothet
5	35	83.3	312	2 S67052	hypothetical prote
6	35	83.3	338	2 T30538	heat shock protein
7	35	83.3	368	2 S75132	sensory transducti
8	35	83.3	420	2 S71199	dnad protein homol
9	35	83.3	491	2 T34499	hypothetical prote
10	35	83.3	510	2 T48977	carboxypeptidase-1
11	35	83.3	925	2 T19361	hypothetical prote
12	35	83.3	1220	2 T06403	resistance complex
13	34	81.0	142	2 B82223	hypothetical prote
14	34	81.0	149	2 T23939	hypothetical prote
15	34	81.0	249	2 S74762	hypothetical prote
16	34	81.0	258	2 E95998	conserved hypothet
17	34	81.0	339	2 B72402	UDP-N-acetylglucos
18	34	81.0	354	2 S52579	prephenate dehydro
19	34	81.0	393	2 A55863	transcriptase - Strept
20	34	81.0	436	2 D70375	gelatinase - Strept
21	34	81.0	437	2 A31142	gelatinase - Strept
22	34	81.0	444	2 T10487	glutamate dehydrog
23	34	81.0	918	2 A55277	hexokinase (EC 2.7
24	33	78.6	115	2 T13519	hypothetical prote
25	33	78.6	211	2 C84162	hypothetical prote
26	33	78.6	287	2 I39689	hypothetical prote
27	33	78.6	310	2 C84701	hypothetical prote
28	33	78.6	332	2 AF3304	oxidoreductase (EC
29	33	78.6	360	2 B71073	probable malate de

30	33	78.6	452	2 F84421	hypothetical prote
31	33	78.6	461	1 T43248	dihydrofolate redu
32	33	78.6	500	1 S22530	carboxypeptidase C
33	33	78.6	507	2 T08337	hypothetical prote
34	33	78.6	529	2 T14947	hypothetical prote
35	33	78.6	565	2 H69113	glycine-tRNA ligase
36	33	78.6	571	2 G89123	protein K07C11.4 (
37	33	78.6	677	2 S15657	transcription fact
38	33	78.6	687	1 A46636	glycine-tRNA ligase
39	33	78.6	701	2 S17196	transcription fact
40	33	78.6	1042	2 S42511	RAG-1 protein - ra
41	33	78.6	1043	2 B33754	recombination-acti
42	33	78.6	1043	2 A33754	recombination-acti
43	33	78.6	1436	2 S67655	probable membrane
44	33	78.6	1456	2 T15961	hypothetical prote
45	33	76.2	129	2 A82547	hypothetical prote

ALIGNMENTS

RESULT 1
T36642
hypothetical protein SCH35.17 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #ext_change 03-Jul-2004
C:Accession: T36642
R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1999
A:Reference number: Z21610
A:Accession: T36642
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-189
A:Cross-references: UNIPROT:Q9X8Y9; EMBL:AL078610; PIR:CA844410.1; GSPDB:GNO0070; SCOEI
C:Genetics:
A:Experimental source: strain A3(2)
A:Gene: SCOEI:SC35.17

Query Match 88.1%; Score 37; DB 2; Length 189;
Best Local Similarity 85.7%; Pred. No. 8.7;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPFE 7
DB 105 KHPQFE 111

RESULT 2
T00484
hypothetical protein At2g35030 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F1913.26
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #ext_change 09-Jul-2004
C:Accession: T00484; G84763
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, April 1998
A:Description: Arabidopsis thaliana chromosome II BAC F1913 genomic sequence.
A:Reference number: Z14160
A:Accession: T00484
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-627 <ROU>
A:Cross-references: UNIPROT:O64766; EMBL:AC004238; NID:G3033373; PID:G3033399
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
Narus, D.; Niernm, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, C.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: G84763
A:Status: preliminary

A/Molecule type: DNA
 A/Residues: 1-627 <STO>
 A/Cross-references: GB:AE002093, NID:G30333399, PIDN:AA12843.1, GSPDB:GN00139
 C/Genetics:
 A/Gene: F1913.26; At2935030
 A/Map position: 2

Query Match 88.1%; Score 37; DB 2; Length 627;
 Best Local Similarity 85.7%; Pred. No. 31;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 KKHPOFEK 7
 Db 590 KSHPOFE 596

RESULT 3
 T10856
 carboxypeptidase C (EC 3.4.16.5) - mung bean (fragment)
 N/Alternate names: serine carboxypeptidase II
 C/Species: Vigna radiata (mung bean)
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
 C/Accession: T10856
 R/Lee, K.; Tan-Wilson, A.L.; Wilson, K.A.
 Submitted to the EMBL Data Library, February 1996
 A/Reference number: Z17185
 A/Accession: T10856
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-101 <LEB>
 A/Cross-references: UNIPROT:Q41689; EMBL:U49382; NID:G1233915; PID:G1233916
 A/Experimental source: strain Jumbo; tissue_type cotyledons; clone pK1054
 C/Keywords: hydrolase, serine carboxypeptidase

Query Match 85.7%; Score 36; DB 2; Length 101;
 Best Local Similarity 75.0%; Pred. No. 6.9;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 KKHPOFEK 8
 Db 82 KKHPOFEK 89

RESULT 4
 G75436
 conserved hypothetical protein - Deinococcus radiodurans (strain R1)
 C/Species: Deinococcus radiodurans
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C/Accession: G75436
 R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; S.; Smith, H.O.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma Science 286, 1571-1577, 1999
 A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A/Reference number: A75250; MUID:20036896; PMID:10567266
 A/Accession: G75436
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-300 <WHI>
 A/Cross-references: UNIPROT:Q9RCV9; GB:AE001960; GB:AE000513; NID:G6458833; PIDN:AAF1067
 A/Experimental source: strain R1
 C/Genetics:
 A/Gene: DR1100
 A/Map position: 1
 C/Superfamily: Escherichia coli ycea protein

Query Match 83.3%; Score 35; DB 2; Length 300;
 Best Local Similarity 62.5%; Pred. No. 34;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 KKHPOFEK 8
 Db 1 KKHPOFEK 8

Db 261 RAHPOFEE 268

RESULT 5
 S67052
 hypothetical protein YOR164c - yeast (Saccharomyces cerevisiae)
 N/Alternate names: hypothetical protein Q3580
 C/Species: Saccharomyces cerevisiae
 C/Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 09-Jul-2004
 C/Accession: S67052
 R/Bordone, R.; Camasses, A.; Medania, A.; Martin, R.P.; Poch, O.; Tarasov, I.A.; Wino submitted to the Protein Sequence Database, July 1996
 A/Reference number: S67032
 A/Accession: S67052
 A/Molecule type: DNA
 A/Residues: 1-312 <BOR>
 A/Cross-references: UNIPROT:Q12125; EMBL:Z75072; NID:G1420404; PID:E252047; PID:G1420405
 A/Experimental source: strain S286C
 C/Genetics:
 A/Gene: MIPS:YOR164c
 A/Cross-references: SGD:S0005690
 A/Map position: 15R

Query Match 83.3%; Score 35; DB 2; Length 312;
 Best Local Similarity 62.5%; Pred. No. 36;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 KKHPOFEK 8
 Db 221 KKHPOFEK 228

RESULT 6
 T30538
 heat shock protein homolog dnaJ - Trypanosoma cruzi
 C/Species: Trypanosoma cruzi
 C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004
 C/Accession: T30538
 R/Bringaud, F.; Vedreine, C.; Cuviellier, A.; Parzy, D.; Baltz, D.; Tetaud, E.; Pays, E. Mol. Biochem. Parasitol. 94, 249-264, 1998
 A/Title: Conserved organization of genes in trypanosomatids.
 A/Reference number: Z16580; MUID:96418771; PMID:9747975
 A/Accession: T30538
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-338 <BRI>
 A/Cross-references: UNIPROT:O76230; EMBL:AF031927; NID:G3452217; PID:G3452219; PIDN:AAC
 C/Genetics:
 A/Gene: dnaJ
 C/Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology <DND>
 F/4-70/Domain: dnaJ amino-terminal homology <DND>

Query Match 83.3%; Score 35; DB 2; Length 338;
 Best Local Similarity 62.5%; Pred. No. 39;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 1 KKHPOFEK 8
 Db 242 KKHPOFEK 249

RESULT 7
 S75132
 sensory transduction system regulatory protein slr2100 - Synechocystis sp. (strain PCC 6
 N/Alternate names: protein slr2100
 C/Species: Synechocystis sp.
 A/Variety: PCC 6803
 C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 12-Jul-2004
 C/Accession: S75132
 R/Kaneko, T.; Sato, S.; Kozami, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1996
 A/Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

```

B.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S75132
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-368 <KAN>
A:Cross-references: EMBL:D90910; GB:AB001339; NID:g1652956; PIDN:BAAL7994.1; PID:g165307
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: response regulator with HD-GYP domain; response regulator homology; sense
C:Keywords: phosphoprotein; transcription regulation
F:13-124/Domain: response regulator homology <RRH>
F:61/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match
Best Local Similarity 83.3%; Score 35; DB 2; Length 368;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKHPOFEK 8
Db 196 KKHPRFK 203

RESULT 8
dhaJ protein homolog atj3 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 28-Oct-1996 #sequence_revision 27-Feb-1997 #text_change 09-Jul-2004
C:Accession: S71199
R:Zhou, R.; Kroczyńska, B.; Miernyk, J.A.
Submitted to the EMBL Data Library, March 1995
A:Description: Atj3, an Arabidopsis thaliana homologue of the Escherichia coli DnaJ.
A:Reference number: S71199
A:Accession: S71199
A:Molecule type: mRNA
A:Residues: 1-420 <ZHO>
A:Cross-references: UNIPROT:Q42530; EMBL:U22340; NID:g1872162; PIDN:AAB49030.1; PID:g727
C:Genetics:
A:Gene: atj3
C:Superfamily: heat shock protein dhaJ, dnaJ amino-terminal homology
F:14-75/Domain: dnaJ amino-terminal homology <DNJ>

Query Match
Best Local Similarity 83.3%; Score 35; DB 2; Length 420;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKHPOFEK 8
Db 260 KKHPOFKR 267

RESULT 9
T34499
hypothetical protein ZK1248.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T34499
R:Lactelle, P.
Submitted to the EMBL Data Library, June 1995
A:Description: The sequence of C. elegans cosmid ZK1248.
A:Reference number: Z21534
A:Accession: T34499
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-491 <LAT>
A:Cross-references: UNIPROT:Q23427; EMBL:U29244; PIDN:AAC71099.1; GSPDB:GN00020; CESP:ZK
A:Experimental source: strain Bristol N2; clone ZK1248
C:Genetics:
A:Gene: CESP:ZK1248.1
A:Map position: 2
A:introns: 34/2; 83/2; 137/3; 230/2; 283/3; 304/3; 364/3; 422/3

Query Match
Best Local Similarity 83.3%; Score 35; DB 2; Length 491;

Query Match
Best Local Similarity 62.5%; Pred. No. 58;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KKHPOFEK 8
Db 591 KKHPOFEK 598

RESULT 10
T48977
carboxypeptidase-like protein F14D17.80 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C:Accession: T48977
R:Jordan, N.; Bangert, S.; Wiedemann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Rudd, S.;
Submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25008
A:Accession: T48977
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-510 <JOR>
A:Cross-references: UNIPROT:Q9LXH4; EMBL:AL353992; GSPDB:GN00061; ATSP:F14D17.80
A:Experimental source: cultivar Columbia; BAC clone F14D17
C:Genetics:
A:Gene: ATSP:F14D17.80
A:introns: 115/2; 171/3; 211/3; 259/3; 310/1; 341/3; 428/1; 472/3
C:Superfamily: serine carboxypeptidase
F:331,421,478/Active site: Ser, Asp, His #status predicted

Query Match
Best Local Similarity 83.3%; Score 35; DB 2; Length 510;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKHPOFEK 8
Db 215 KKHPOFKV 222

RESULT 11
T19361
hypothetical protein C17G1.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T19361
R:White, S.
Submitted to the EMBL Data Library, August 1996
A:Reference number: Z19114
A:Accession: T19361
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-925 <WIL>
A:Cross-references: UNIPROT:Q93238; EMBL:Z78415; PIDN:CAB01670.1; GSPDB:GN00028; CESP:CL
A:Experimental source: clone C17G1
C:Genetics:
A:Gene: CESP:C17G1.4
A:Map position: X
A:introns: 40/3; 98/2; 295/1; 443/3; 590/3; 619/3; 691/1; 810/1; 868/3
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology,

Query Match
Best Local Similarity 83.3%; Score 35; DB 2; Length 925;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KKHPOFEK 8
Db 591 KKHPOFEK 598

RESULT 12
T06403
resistance complex protein I2C-1 - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 09-Jul-2004

```

C/Accession: T06403
 R:Ori, N.; Eshed, Y.; Paran, I.; Presting, G.; Aviv, D.; Tanksley, S.; Zamir, D.; Fluhr, Plant Cell 9, 521-532, 1997
 A>Title: The 12C family from the wilt disease resistance locus 12 belongs to the nucleos A;Reference number: 215652; MUID:97290204; PMID:9144960
 A/Accession: T06403
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-1220 <ORI>
 A/Cross-references: UNIPROT:Q24015; EMBL:AF004878; NID:92258314; PIDN:AB63274.1; PID:92 C/Genetics:
 A:Gene: 12C-1
 A:Map position: 11
 C/Function:
 A>Description: confers resistance against Fusarium oxysporum
 C/Superfamily: disease resistance protein RPS2; leucine-rich alpha-2-glycoprotein repeat

Query Match 83.3%; Score 35; DB 2; Length 1220;
 Best Local Similarity 62.5%; Pred. No. 1.6e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKHPOFEK 8
 DB 362 KKHPEFE 369

RESULT 13
 B82223
 hypothetical protein VC1253 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
 C/Species: Vibrio cholerae
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
 C/Accession: B82223
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Charadon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bacs, S.; Qin, H.; Dragoti, I.; Sellers, F. 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A/Reference number: A82035; MUID:20406833; PMID:10952301
 A/Accession: B82223
 A>Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-142 <HEI>
 A/Cross-references: UNIPROT:Q9SKJ3; GB:AE004204; GB:AE003852; NID:99655729; PIDN:AAF9441
 A/Experimental source: serogroup O1; strain N16961; biotype El Tor
 C/Genetics:
 A:Gene: VC1253
 A/Map position: 1

Query Match 81.0%; Score 34; DB 2; Length 142;
 Best Local Similarity 62.5%; Pred. No. 24;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKHPOFEK 8
 DB 86 KKHPOFRR 93

RESULT 14
 T23939
 hypothetical protein R05H5.3 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C/Accession: T23939
 R:McMurray, A.
 submitted to the EMBL Data Library, March 1995
 A/Reference number: Z19821
 A/Accession: T23939
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-149 <WII>
 A/Cross-references: UNIPROT:Q21763; EMBL:Z48795; PIDN:CAA88726.1; GSPDB:GNO0020; CESP:RC A/Experimental source: clone R05H5
 C/Genetics:

A/Gene: CESP:R05H5.3
 A/Map position: 2
 A/Introns: 27/3; 98/2
 C/Superfamily: Caenorhabditis elegans hypothetical protein C35B1.5

Query Match 81.0%; Score 34; DB 2; Length 149;
 Best Local Similarity 71.4%; Pred. No. 25;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKHPOFE 7
 DB 57 KKHPEFE 63

RESULT 15
 S74762
 hypothetical protein alr1619 - Synechocystis sp. (strain PCC 6803)
 C/Species: Synechocystis sp.
 A/Variety: PCC 6803
 C/Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 C/Accession: S74762
 R:Kaneke, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda DNA Res. 3, 109-136, 1995
 A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 A/Reference number: S74322; MUID:97061201; PMID:8905231
 A/Accession: S74762
 A>Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-249 <KAN>
 A/Cross-references: UNIPROT:P72897; EMBL:D90901; GB:AB001339; NID:91651897
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 81.0%; Score 34; DB 2; Length 249;
 Best Local Similarity 71.4%; Pred. No. 44;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKHPOFE 7
 DB 220 KKHPEFE 226

Search completed: March 2, 2005, 12:28:52
 Job time : 8.11382 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 29.7236 Seconds

(without alignments)
137.824 Million cell updates/sec

Title: SEQ5

Perfect score: 42

Sequence: 1 kxhpgef 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Uniprot 03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	209	2	06LFL4
2	39	92.9	323	2	08LFT5
3	38	90.5	410	2	073B87
4	37	88.1	182	2	088XG7
5	37	88.1	189	2	09X8Y9
6	37	88.1	335	2	07V101
7	37	88.1	439	1	1DGI_DROVA
8	37	88.1	627	2	064766
9	37	88.1	704	2	08CJD2
10	37	88.1	743	2	08CH90
11	37	88.1	2932	2	0814T6
12	37	88.1	3409	2	07S127
13	36	85.7	101	2	041689
14	36	85.7	162	2	08E887
15	36	85.7	257	2	09VPS1
16	36	85.7	714	1	06P8H2
17	36	85.7	715	1	08B8R3
18	36	85.7	715	2	08B8R3
19	36	85.7	715	2	08C395
20	35	83.3	91	2	08RTY3
21	35	83.3	223	2	08B636
22	35	83.3	300	1	YB00_DEIRA
23	35	83.3	312	2	012125
24	35	83.3	319	2	0814Y8
25	35	83.3	338	2	076230
26	35	83.3	338	2	09B1X8
27	35	83.3	368	2	P73328
28	35	83.3	420	2	042530
29	35	83.3	510	2	091XK4
30	35	83.3	510	2	091XK4
31	35	83.3	515	2	023427

32	35	83.3	532	2	036160	036160 vibrio chol
33	35	83.3	800	2	08W759	08W759 bacterioph
34	35	83.3	925	2	093238	093238 caenorhabd
35	35	83.3	989	2	064055	064055 bacteroides
36	35	83.3	1220	2	024015	024015 lycopersico
37	36	81.0	48	2	048504	048504 lactococcus
38	34	81.0	70	2	06N0C1	06N0C1 magnetospi
39	34	81.0	89	1	RS17_LEPIN	09X427 leptospira
40	34	81.0	89	2	072NHO	072NHO leptospira
41	34	81.0	107	2	06Z315	06Z315 oryza sativ
42	34	81.0	140	2	08GHJ5	08GHJ5 thermus the
43	34	81.0	140	2	072HFI	072HFI thermus the
44	34	81.0	142	2	09KSK3	09KSK3 vibrio chol
45	34	81.0	149	2	021763	021763 caenorhabd

ALIGNMENTS

RESULT 1

ID	Q6LFL4	PRELIMINARY;	PRT;	209 AA.
AC	Q6LFL4;			
DT	05-JUL-2004 (TRMBLrel. 27, Created)			
DT	05-JUL-2004 (TRMBLrel. 27, Last sequence update)			
DT	05-JUL-2004 (TRMBLrel. 27, Last annotation update)			
DE	Hypothetical protein.			
GN	ORFNames=MAL6P1.55, PPF0255c;			
OS	Plasmodium falciparum (Isolate 3D7).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=36329;			
RX	SEQUENCE FROM N.A.			
RX	PubMed=12368867; DOI=10.1038/nature01095;			
RA	Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,			
RA	Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,			
RA	Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,			
RA	Chillingworth T., Christodoulou Z., Clark L., Clark R., Corto C.,			
RA	Cronin A., Davies R., Davies P., Dear P., Dearden F., Doggett J.,			
RA	Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,			
RA	Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,			
RA	Humphrey S., Jagels K., James D., Johnson D., Kethornu A., Knight A.,			
RA	Kontorov B., Keyes S., Larke N., Lawson D., Leonard N., Line A.,			
RA	Maddison M., McLean J., Mooney P., Moule S., Murphy L., Oliver K.,			
RA	Ormond D., Price C., Quail M.A., Rabbinowitch E., Rajandream M.A.,			
RA	Rutter S., Rutherford K.M., Sanders M., Simmonds M., Seeger K.,			
RA	Sharp S., Smith R., Squares R., Squares S., Stevens K., Taylor K.,			
RA	Tivee A., Unwin L., Whitehead S., Woodward J., Sulston J.E., Craig A.,			
RA	Newbold C., Barrrell B.G.			
RT	"Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."			
RL	Nature 419:527-531 (2002).			
RN	(2)			
RP	SEQUENCE FROM N.A.			
RA	Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,			
RA	Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,			
RA	Ormond D., Sanders M., Hayes R., Hall S., Quail M.A., Barrrell B.G.			
RL	Submitted (Mar-2004) to the EMBL/Genbank/DBS databases.			
DR	EMBL; CR382398; CAG25222.1; -			
DR	GO; GO:0003989; F:acylphosphatase activity; IEA.			
DR	InterPro; IPR001792; Acylphosphatase.			
DR	Pfam; PF00708; Acylphosphatase; 1.			
KW	Hypothetical protein.			
SQ	SEQUENCE 209 AA; 25373 MW; F1682AB63776B3A2 CRC64;			

Query Match 92.9%; Score 39; DB 2; Length 209;

Best Local Similarity 75.0%; Pred. No. 14;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY	1 KXHPGEK 8
DB	44 KXHPGEK 51

RESULT 2

08LTT5 PRELIMINARY; PRT; 323 AA.
 AC 08LTT5;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE 5' to 3' exonuclease-like protein.
 OS Vibrio parahaemolyticus.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;
 OC T7-like viruses.
 OX NCBI_TaxID=194802;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22667917; PubMed=12781722;
 RA Hardies S.C., Comeau A.M., Serwer P., Suttle C.A.;
 RT "The complete sequence of marine bacteriophage VP262 infecting vibrio parahaemolyticus indicates that an ancestral component of a T7 viral supergroup is widespread in the marine environment."
 RL Virology 310:359-371(2003).
 RL EMBL; AY095314; AAM28378.1;
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0004527; F:exonuclease activity; IEA.
 DR InterPro; IPR008918; 5_3_exo_C.
 DR InterPro; IPR000513; Exo_N_I.
 KW Exonuclease.
 SQ SEQUENCE 323 AA; 37725 MW; 81CB297839533951 CRC64;

Query Match

Best Local Similarity 92.9%; Score 39; DB 2; Length 323;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KXHPQFEK 8
 Db 76 KKHPRFEK 83

RESULT 3

073E87 PRELIMINARY; PRT; 410 AA.
 AC 073E87;
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
 DE Arginine deiminase (EC 3.5.3.6).
 GN Name=arica; OrderedLocustNames=BCE0472;
 OS Bacillus cereus (strain ATCC 10987).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=222523;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX PubMed=14960714; DOI=10.1093/nar/gkh258;
 RA Raasbo D.A., Ravel J., Oekstad O.A., Helgason E., Ger R.Z., Jiang L., Shores K.A., Foutz D.E., Tourassis N.J., Anguino J.V., Kolonay J.F., Nelson W.C., Kolstoe A.-B., Fraser C.M., Reed T.D.;
 RT "The genome sequence of Bacillus cereus ATCC 10987 reveals metabolic adaptations and a large plasmid related to Bacillus anthracis pXO1."
 RL Nucleic Acids Res. 32:977-988(2004).
 RL EMBL; AE017265; AAS39407.1;
 DR TIGR; BCE0472;
 DR GO; GO:0016990; F:arginine deiminase activity; IEA.
 DR GO; GO:0016787; F:hydrolyase activity; IEA.
 DR GO; GO:0006527; P:arginine catabolism; IEA.
 DR InterPro; IPR003198; Amidino_trans.
 DR InterPro; IPR003876; Arg_deiminase.
 DR Pfam; PF02274; Amidinoctransf; 1.
 DR PRINTS; PR01466; ARGDEIMINASE.
 DR TIGRPFMS; TIGR01078; arca; 1.
 KW Complete proteome; Hydrolyase.
 SQ SEQUENCE 410 AA; 46938 MW; 8D7846C8D206579A CRC64;

Query Match 90.5%; Score 38; DB 2; Length 410;
 Best Local Similarity 75.0%; Pred. No. 47;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KXHPQFEK 8
 Db 192 KKHPRFEK 199

RESULT 4

088XG7 PRELIMINARY; PRT; 182 AA.
 AC 088XG7;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Hypothetical protein lp_1238.
 GN OrderedLocustNames=lp_1238;
 OS Lactobacillus plantarum.
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
 OC Lactobacillus.
 OX NCBI_TaxID=1590;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCIMB 8826 / WCFS1;
 RX MEDLINE=22480296; PubMed=12566566; DOI=10.1073/pnas.0337704100;
 RA Kleerebezem M., Boekhorst J., van Kramenburg R., Molenaar D., Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandbrink H.M., Rieters M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A., Roffler S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ursing B., De Vos W.M., Sijzen R.J.;
 RT "Complete genome sequence of Lactobacillus plantarum WCFS1."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
 DR EMBL; AL935255; CAD63746.1;
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 182 AA; 21722 MW; 0C1241611A43D8CF CRC64;

Query Match

Best Local Similarity 88.1%; Score 37; DB 2; Length 182;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KXHPQFEK 7
 Db 35 KXHPQFEK 41

RESULT 5

09X8Y9 PRELIMINARY; PRT; 189 AA.
 AC 09X8Y9;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Putative lipoprotein.
 GN ORFNames=CH33.17;
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteriota; Actinobacteridae; Actinomycetales;
 OC Streptomycetales; Streptomycetaceae; Streptomycetes.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;
 RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L., Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H., Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M., Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S., Huang C.-H., Kieiser J., Larke L., Murphy L.D., Oliver K., O'Neill S., Rabinowitsch E., Rajandream M.A., Rutherford K.M., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K., Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J., Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2)."
 RL Nature 417:141-147(2002).

DR EMBL AL939117; CAB44410.1; -.
DR PIR T36642; T36642.
KW Complete proteome; Lipoprotein.
SQ SEQUENCE 189 AA; 19772 MW; 82D231E01CAC3B57 CRC64;
Query Match 88.1%; Score 37; DB 2; Length 189;
Best Local Similarity 85.7%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKHPOFE 7
DB 105 KKHPOFE 111
RESULT 6
QY101 PRELIMINARY; PRT; 335 AA.
ID QY101
AC QY101
DT 01-OCT-2003 (TREMblrel. 25, Created)
DT 01-OCT-2003 (TREMblrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)
DE Biotin synthase (EC 2.8.1.6).
GN Name=Biot; OrderedLocustNames=PMW1093;
OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).
OC Bacteria; Cyanobacteria; Prochlorales; Prochlorococcales;
OC Prochlorococcus.
OX NCBI_TaxID=59919;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22825698; PubMed=12917642; DOI=10.1038/nature01947;
RA Rood G., Larimer F.W., Lamerding J.E., Malfatti S., Chain P.,
RA Ahlren N.A., Arelano A., Coleman M., Hauser L., Hess W.R.,
RA Johnson Z.I., Land M.L., Lindell D., Post A.F., Regala W., Shah M.,
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
RA Webb E.A., Zinner E.R., Chisholm S.W.;
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
RT niche differentiation."
RT Nature 424:1042-1047(2003).
DR EMBL BX572092; CAB19552.1; -.
DR HSP; P12986; IR30.
DR GO; GO:0004076; Fibrin synthase activity; IEA.
DR GO; GO:0005506; Fibrin ion binding; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009102; P:biotin biosynthesis; IEA.
DR InterPro; IPR010722; BATS.
DR InterPro; IPR002684; Biotin_synth.
DR InterPro; IPR006358; Elp3/MiAB/NiB.
DR Pfam; PF06968; BATS; 1.
DR Pfam; PF04055; Radical_SAM.
DR SMART; SM00729; Elp3; 1.
DR TIGRFAMs; TIGR00433; bIob; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 335 AA; 37719 MW; EB9F3FB56044E04D CRC64;
Query Match 88.1%; Score 37; DB 2; Length 335;
Best Local Similarity 85.7%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKHPOFE 7
DB 79 KKHPOFE 85
RESULT 7
ID IDG1 DROYA STANDARD; PRT; 439 AA.
AC O8MX40;
DT 29-MAR-2004 (rel. 43, Created)
DT 29-MAR-2004 (rel. 43, Last sequence update)
DT 25-OCT-2004 (rel. 45, Last annotation update)
DE Chitinase-like protein idgfl precursor (imaginal disk growth factor
protein 1).
GN Name=idgfl;
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=#14021-0261.0;
RX MEDLINE=22226511; PubMed=12242232;
RA Zurovcova M., Ayala F.J.;
RT "Polymorphism patterns in two tightly linked developmental genes,
RT idgf1 and idgf3, of Drosophila melanogaster."
RT Genetics 162:177-188(2002).
CC - FUNCTION: Cooperates with insulin-like peptides to stimulate the
CC proliferation, polarization and motility of imaginal disk cells.
CC May act by stabilizing the binding of insulin-like peptides to its
CC receptor through a simultaneous interaction with both molecules to
CC form a multiprotein signaling complex (By similarity).
CC - SUBCELLULAR LOCATION: Secreted. It is probably transported to
CC target tissues via hemolymph (By similarity).
CC - PTM: Glycosylated (By similarity).
CC - MISCELLANEOUS: Lacks the typical Glu active site in position 150
CC that is replaced by a Gln residue, preventing the hydrolase
CC activity. Its precise function remains unclear.
CC - SIMILARITY: Belongs to the glycosyl hydrolase 18 family. IDGP
CC subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL AF394712; AAM69644.1; -.
DR HSP; O96665; LUND.
DR Flybase; FBgn0044136; Dyak1Idgf1.
DR InterPro; IPR001223; Glyco_hydro_18.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD0004711; Chitinase_II; 1.
DR SMART; SM00636; Glyco_18; 1.
KW Developmental protein; Glycoprotein; Signal.
FT SIGNAL 1 20
FT CHAIN 21 439
FT FT 26 53
FT DISULFID 340 423
FT FT 122 122
FT CARBOHYD 218 218
FT CARBOHYD 346 346
SQ SEQUENCE 439 AA; 49244 MW; EDE16BFD82A1B9E CRC64;
Query Match 88.1%; Score 37; DB 1; Length 439;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKHPOFE 7
DB 227 KKHPOFE 233
RESULT 8
ID O64766 PRELIMINARY; PRT; 627 AA.
AC O64766;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)
DE Hypothetical protein At2g35030.
GN Name=At2g35030;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

GN Name=idgfl;
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=#14021-0261.0;
RX MEDLINE=22226511; PubMed=12242232;
RA Zurovcova M., Ayala F.J.;
RT "Polymorphism patterns in two tightly linked developmental genes,
RT idgf1 and idgf3, of Drosophila melanogaster."
RT Genetics 162:177-188(2002).
CC - FUNCTION: Cooperates with insulin-like peptides to stimulate the
CC proliferation, polarization and motility of imaginal disk cells.
CC May act by stabilizing the binding of insulin-like peptides to its
CC receptor through a simultaneous interaction with both molecules to
CC form a multiprotein signaling complex (By similarity).
CC - SUBCELLULAR LOCATION: Secreted. It is probably transported to
CC target tissues via hemolymph (By similarity).
CC - PTM: Glycosylated (By similarity).
CC - MISCELLANEOUS: Lacks the typical Glu active site in position 150
CC that is replaced by a Gln residue, preventing the hydrolase
CC activity. Its precise function remains unclear.
CC - SIMILARITY: Belongs to the glycosyl hydrolase 18 family. IDGP
CC subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL AF394712; AAM69644.1; -.
DR HSP; O96665; LUND.
DR Flybase; FBgn0044136; Dyak1Idgf1.
DR InterPro; IPR001223; Glyco_hydro_18.
DR Pfam; PF00704; Glyco_hydro_18; 1.
DR ProDom; PD0004711; Chitinase_II; 1.
DR SMART; SM00636; Glyco_18; 1.
KW Developmental protein; Glycoprotein; Signal.
FT SIGNAL 1 20
FT CHAIN 21 439
FT FT 26 53
FT DISULFID 340 423
FT FT 122 122
FT CARBOHYD 218 218
FT CARBOHYD 346 346
SQ SEQUENCE 439 AA; 49244 MW; EDE16BFD82A1B9E CRC64;
Query Match 88.1%; Score 37; DB 1; Length 439;
Best Local Similarity 85.7%; Pred. No. 80;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 KKHPOFE 7
DB 227 KKHPOFE 233
RESULT 8
ID O64766 PRELIMINARY; PRT; 627 AA.
AC O64766;
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-MAR-2004 (TREMblrel. 26, Last annotation update)
DE Hypothetical protein At2g35030.
GN Name=At2g35030;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

OC  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC  eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX  NCBI_TaxID=3702;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
RA  Symes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
RA  Somerville C.R., Venter J.C.;
RL  Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RA  Town C.D., Kaul S.;
RL  Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AC004238; AAC12843.1; -.
DR  PIR; T00484; T00484.
DR  InterPro; IPR002885; PPR.
DR  InterPro; IPR008940; Premyl trans.
DR  InterPro; IPR008941; TPR-like.
DR  Pfam; PF01535; PPR; 12.
DR  TIGRPFAM; TIGR00756; PPR; 10.
KW  Hypothetical protein.
SQ  SEQUENCE 627 AA; 71409 MW; 27BDA2BC168AF949 CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 2; Length 627;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY  1 KKHPOFE 7
DB  590 KSHPOFE 596

RESULT 9
OQ8CJD2  PRELIMINARY; PRT; 704 AA.
ID  OQ8CJD2;
AC  OQ8CJD2;
DT  01-MAR-2003 (TReMBLrel. 23, Created)
DT  01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT  01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE  Guanylyl cyclase alpha 1 subunit.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Nakamura I., Suzuki N.;
RL  Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AB096020; BAC24016.1; -.
DR  HSSP; P30803; IAZS.
DR  GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR  GO; GO:0016829; F:lyase activity; IEA.
DR  GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR  InterPro; IPR001054; G_cyclase.
DR  Pfam; PF00211; Guanylate_cyc; 1.
DR  SMART; SM00044; CYCC; 1.
DR  PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR  PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW  Lyase.
SQ  SEQUENCE 704 AA; 79188 MW; F73FC97B68544E6A CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 2; Length 704;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY  2 XHPOFEK 8
DB  698 SHPOFEK 704

RESULT 10
OQ8CH90  PRELIMINARY; PRT; 743 AA.

```

```

AC  OQ8CH90;
DT  01-MAR-2003 (TReMBLrel. 23, Created)
DT  01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT  01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE  Soluble guanylyl cyclase alpha 2 subunit E219G mutant.
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Tissue=Kidney;
RA  Nakamura I., Yao Y., Suzuki N.;
RL  Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AB097860; BAC44887.1; -.
DR  HSSP; P30803; IAZS.
DR  GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR  GO; GO:0016829; F:lyase activity; IEA.
DR  GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR  InterPro; IPR001054; G_cyclase.
DR  InterPro; IPR009080; tRNAsyn_1a_bind.
DR  Pfam; PF00211; Guanylate_cyc; 1.
DR  SMART; SM00044; CYCC; 1.
DR  PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR  PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW  Lyase.
SQ  SEQUENCE 743 AA; 83251 MW; ACF5C53E0982813A CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 2; Length 743;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY  2 XHPOFEK 8
DB  737 SHPOFEK 743

RESULT 11
OQ8I4T6  PRELIMINARY; PRT; 2932 AA.
ID  OQ8I4T6;
AC  OQ8I4T6;
DT  01-MAR-2003 (TReMBLrel. 23, Created)
DT  01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT  01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE  Hypothetical protein.
GN  ORPNames=PF12399c.
OS  Plasmodium falciparum (Isolate 3D7).
OC  Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX  NCBI_TaxID=36329;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  MEDLINE=2255705; PubMed=12368864; DOI=10.1038/nature01097;
RA  Gardner M.J., Hall N., Fung B., White O., Berriman M., Hyman R.W.,
RA  Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
RA  Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
RA  Chan M.S., Nene V., Shallow S.J., Suh B., Peterson J., Anjiloli S.,
RA  Pertes M., Allen J., Selengut J., Haft D., Mether M.W., Vaidya A.B.,
RA  Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
RA  McPadden G.I., Cummings J.M., Subramanian G.M., Mungall C.,
RA  Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
RA  Fraser C.M., Barrett B.;
RT  "Genome sequence of the human malaria parasite Plasmodium
RT  falciparum."
RL  Nature 419:498-511 (2002).
RN  [2]
RP  SEQUENCE FROM N.A.
RA  Hyman R.W., Fung B., Conway A., Kurdi O., Mao J., Miranda M.,
RA  Nakao B., Rowley D., Tamaki T., Wang F., Wang R.W.;
RL  Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AE014852; AAN35562.1; -.
KW  Hypothetical protein.
SQ  SEQUENCE 2932 AA; 348674 MW; BA806DBC03CC11C8 CRC64;

```

Query Match 88.1%; Score 37; DB 2; Length 2932;
 Best Local Similarity 75.0%; Pred. No. 6.4e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KXHPQFEK 8
 DB 1499 KXHPNFEK 1506

RESULT 12
 ID Q7S127 PRELIMINARY; PRT; 3409 AA.
 AC Q7S127;
 DT 01-MAR-2004 (TrEMBLrel. 26, Created)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
 DE Hypothetical protein.
 GN Name=NCU00625.1;
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OR74A;
 RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
 Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Puccell S., Rehm B.,
 Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
 Qui D., Iankilev P., Pedersen D., Nelson M., Washburne M.,
 Seltremitdoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
 Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
 Roy A., Foldy K., Naylor J., Thomann N., Barrett R., Gierre S.,
 Kamal M., Kamysheva M., Mauceli E., Bielke C., Rudd S., Frishman D.,
 Kryatova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,
 Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmann S.A.,
 Desouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
 Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
 Natvig D.O., Alex L.A., Manhaupt G., Ebbole D.J., Freitag M.,
 Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.;
 RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
 RL Nature 0:0-0(2003).
 CC -1- CAUTION: The sequence shown here is derived from an
 EMBL/GenBank/DDJ whole genome shotgun (WGS) entry which is
 preliminary data
 CC EMBL: AABX01000001; EAA35529.1; -.
 DR HSSP; P48736; IE8Y.
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0005786; C:signal recognition particle (sensu Eukaryota); IEA.
 DR GO; GO:0005525; F:GTP binding; IEA.
 DR GO; GO:0016773; F:phosphotransferase activity, alcohol group . . .; IEA.
 DR GO; GO:0003723; F:RNA binding; IEA.
 DR GO; GO:0006886; P:intracellular protein transport; IEA.
 DR GO; GO:0006814; P:SRP-dependent cotranslational protein-membr. . .; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR003151; FAT.
 DR InterPro; IPR003152; FAT.
 DR InterPro; IPR011009; Kinase_like.
 DR InterPro; IPR011012; Longin_like.
 DR InterPro; IPR000403; P13_P14_kinase.
 DR InterPro; IPR008940; Prenyl_trans.
 DR InterPro; IPR007222; SRP-alpha_N.
 DR InterPro; IPR000897; SRP54.
 DR Pfam; PF02259; FAT; 1.
 DR Pfam; PF02260; FATC; 1.
 DR Pfam; PF00454; P13_P14_kinase; 1.
 DR Pfam; PF00486; SRP-alpha_N; 1.
 DR Pfam; PF00448; SRP54; 1.
 DR ProDom; PD000819; SRP54; 1.
 DR PROSITE; PS50290; P13_4_KINASE_3; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 3409 AA; 379497 MW; D6548734032D01BC CRC64;

Query Match 88.1%; Score 37; DB 2; Length 3409;
 Best Local Similarity 75.0%; Pred. No. 7.6e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFEK 8
 DB 2403 KTHPQFEK 2410

RESULT 13
 ID Q41689 PRELIMINARY; PRT; 101 AA.
 AC Q41689;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE Serine carboxypeptidase (Fragment).
 OS Vigna radiata.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosoid I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
 OX NCBI_TaxID=157791;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Jumbo; TISSUE=Cotyledons;
 RA Lee K., Tan-Wilson A.L., Wilson K.A.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DDJ databases.
 DR EMBL; U9382; AAA92062.1; -.
 DR PIR; T10856; T10856.
 DR HSSP; P08819; 1MHT.
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR GO; GO:001685; F:serine carboxypeptidase activity; IEA.
 DR GO; GO:0005508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR001563; Peptidase_S10.
 DR InterPro; IPR000379; Ser_estr.
 DR Pfam; PF00450; Peptidase_S10; 1.
 DR PRINTS; PR00724; CRBOXYPASRC.
 DR ProDom; PD001189; Peptidase_S10; 1.
 DR PROSITE; PS00131; CARBOXYPEPT_SER_SER; 1.
 KW Carboxypeptidase.
 FT NON TER 1 101
 FT NON TER 1 101
 SQ SEQUENCE 101 AA; 11467 MW; 40B371DBB2D53818 CRC64;
 Query Match 85.7%; Score 36; DB 2; Length 101;
 Best Local Similarity 75.0%; Pred. No. 25;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KXHPQFEK 8
 DB 82 KHPQFEK 89

RESULT 14
 ID Q8E887 PRELIMINARY; PRT; 162 AA.
 AC Q8E887;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein SOA00033.
 GN OrdereddictusNames=SOA00033;
 OS Shewanella oneidensis.
 OG Shewanella megaplasmid.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Alteromonadales;
 OC Shewanellaceae; Shewanella.
 OX NCBI_TaxID=70863;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MR-1;
 RX MEDLINE=22297686; PubMed=12368813; DOI=10.1038/nbt749;
 RA Heidelberg J.F., Paulsen I.T., Nelson K.E., Gaidos E.J., Nelson W.C.,
 Read T.D., Eisen J.A., Seshadri R., Ward N.L., Methe B.A.,

```

RA Claydon R.A., Meyer T., Tsaplin A., Scott J., Beaman M.J.,
RA Brinkac L.M., Daugherty S.C., DeBoy R.T., Dodson R.J., Durkin A.S.,
RA Haft D.H., Kolony J.F., Madupu R., Peterson J.D., Umayam L.A.,
RA White O., Wolf A.M., Vanathavan J.J., Weidman J.F., Impraim M.,
RA Lee K., Berry K.V., Lee C., Mueller J., Khouri H.M., Gill J.,
RA Utecherback T.R., McDonald L.A., Feldblyum T.V., Smith H.O.,
RA Venter J.C., Neilson K.H., Fraser C.M.;
RT "Genome sequence of the dissimilatory metal ion-reducing bacterium
RT Shewanella oneidensis."
RT Nat. Biotechnol. 20:1118-1123(2002).
DR EMBL: AB015911; AN52969.1; -.
DR TIGR: SOA0033; -.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0016567; P:protein ubiquitination; IEA.
DR InterPro: IPR002714; VHL.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 162 AA; 18141 MW; F8F94B07061CB7C5 CRC64;

Query Match      85.7%; Score 36; DB 2; Length 162;
Best Local Similarity 62.5%; Pred. No. 43;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 KKHPOFEK 8
DB      144 KKHPOFEK 151

RESULT 15
QYVPS1 PRELIMINARY; PRT; 257 AA.
AC QYVPS1;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DB CG13686-PA.
DE CG13686-PA.
GN Name:lectin-21Cb; ORFNames:CG13686;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazek R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaborel G.L.,
RA Abertl J.P., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Beres P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrelia S., Fleischmann W.,
RA Fjoele A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Heston D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Urali M., Kalush F., Karpen G.H., Ke Z., Kenison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai X.,
RA Lesko P., Lei Y., Levitsky A.A., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Murphy L., Muzny D.M., Nelson D.L.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

```

```

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodagel, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=2242605; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskaas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RT Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskaas R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomic perspective."
RT Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
[4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Mitra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Bernan B.P.,
RA Belencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review."
RT Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
[5]
RP SEQUENCE FROM N.A.
RX Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RL [6]
RP SEQUENCE FROM N.A.
RG FlyBase;
RG FlyBase;
RL EMBL: AE003588; AAF51471.1; -.
DR HSSP: P05452; 1HTN.
DR FlyBase; FBGN0040106; lectin-21Cb.
DR GO: GO:0005529; F:sugar binding; IEA.
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00059; Lectin_C; 1.
DR SMART: SMO0034; LECTIN; 1.
DR PROSITE: PS50041; C-TYPE LECTIN 2; 1.
SQ SEQUENCE 257 AA; 30032 MW; BA21A7E55686CA39 CRC64;

Query Match      85.7%; Score 36; DB 2; Length 257;
Best Local Similarity 62.5%; Pred. No. 71;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 KKHPOFEK 8
DB      135 KKHPOFEK 142

Search completed: March 2, 2005, 12:44:16
Job time : 32.7236 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 36.0325 Seconds
(without alignments)
85.863 Million cell updates/sec

Title: SEQ5

Perfect score: 42

Sequence: i kxhpfek 8

Scoring table:

BL0SUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 366760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq16Dec04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	598	8	ABM83512 Human dia
2	37	88.1	8	2	AAW59212 Streptavi
3	37	88.1	8	4	AAJ35433 Epitope P
4	37	88.1	8	4	AAJ35442 Nascent P
5	37	88.1	8	4	AAJ35442 Strept-Tag
6	37	88.1	8	5	AAJ35442 Strept-Tag
7	37	88.1	8	5	AAJ35442 Strept-Tag
8	37	88.1	8	5	AAJ35442 Strept-Tag
9	37	88.1	8	5	AAJ35442 Strept-Tag
10	37	88.1	8	6	AAJ35442 Strept-Tag
11	37	88.1	8	6	AAJ35442 Strept-Tag
12	37	88.1	8	6	AAJ35442 Strept-Tag
13	37	88.1	8	7	AAJ35442 Strept-Tag
14	37	88.1	8	7	AAJ35442 Strept-Tag
15	37	88.1	8	7	AAJ35442 Strept-Tag
16	37	88.1	8	7	AAJ35442 Strept-Tag
17	37	88.1	8	7	AAJ35442 Strept-Tag
18	37	88.1	8	8	AAJ35442 Strept-Tag
19	37	88.1	8	8	AAJ35442 Strept-Tag
20	37	88.1	8	8	AAJ35442 Strept-Tag
21	37	88.1	8	8	AAJ35442 Strept-Tag
22	37	88.1	8	8	AAJ35442 Strept-Tag
23	37	88.1	8	8	AAJ35442 Strept-Tag
24	37	88.1	8	8	AAJ35442 Strept-Tag
25	37	88.1	8	8	AAJ35442 Strept-Tag

26	37	88.1	10	8	ADJ11064 Strept-tag
27	37	88.1	10	8	ADJ11064 Strept-tag
28	37	88.1	10	8	ADJ11064 Strept-tag
29	37	88.1	11	6	AAJ35433 Epitope P
30	37	88.1	11	8	AAJ35433 Epitope P
31	37	88.1	11	8	AAJ35433 Epitope P
32	37	88.1	11	8	AAJ35433 Epitope P
33	37	88.1	11	8	AAJ35433 Epitope P
34	37	88.1	11	8	AAJ35433 Epitope P
35	37	88.1	11	8	AAJ35433 Epitope P
36	37	88.1	11	8	AAJ35433 Epitope P
37	37	88.1	11	8	AAJ35433 Epitope P
38	37	88.1	11	8	AAJ35433 Epitope P
39	37	88.1	11	8	AAJ35433 Epitope P
40	37	88.1	11	8	AAJ35433 Epitope P
41	37	88.1	11	8	AAJ35433 Epitope P
42	37	88.1	11	8	AAJ35433 Epitope P
43	37	88.1	11	8	AAJ35433 Epitope P
44	37	88.1	11	8	AAJ35433 Epitope P
45	37	88.1	11	8	AAJ35433 Epitope P

ALIGNMENTS

RESULT 1	ABM83512	standard; protein; 598 AA.
ID	ABM83512	standard; protein; 598 AA.
AC	ABM83512;	
XX	18-NOV-2004	(first entry)
DT	18-NOV-2004	(first entry)
XX	Human diagnostic and therapeutic protein SEQ ID NO:3761.	
DE	Human diagnostic and therapeutic protein SEQ ID NO:3761.	
XX	gene therapy; human diagnostic and therapeutic polynucleotide; dittp.	
KW	gene therapy; human diagnostic and therapeutic polynucleotide; dittp.	
XX	Homo sapiens.	
OS	Homo sapiens.	
XX	WO2004023973-A2.	
PN	WO2004023973-A2.	
XX	25-MAR-2004.	
PD	25-MAR-2004.	
XX	12-SEP-2003; 2003WO-US028227.	
PF	12-SEP-2003; 2003WO-US028227.	
XX	12-SEP-2002; 2002US-0410259P.	
PR	12-SEP-2002; 2002US-0410259P.	
XX	12-SEP-2002; 2002US-0410259P.	
XX	(INCY-) INCYTE CORP.	
PA	(INCY-) INCYTE CORP.	
PI	Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F,	
XX	Schmidt JP, Wright RJ, Bruns CM, Marjanovic MM, Shen F,	
PI	Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Rider LV,	
XX	Harthorne TA, Suchorolski MT, Altus CM, Pitts SJ, Rider LV,	
PI	Mooney EM, Deleage AM, Parnesar IS, Banville SC, Reddy TP,	
XX	Mooney EM, Deleage AM, Parnesar IS, Banville SC, Reddy TP,	
PI	Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerecht EH,	
XX	Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerecht EH,	
PI	Perla CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LB,	
XX	Perla CH, Anderson SB, Rioux P, Shen EJ, Wu MC, Stuve LB,	
PI	Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton BS,	
XX	Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kirton BS,	
PI	Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gierzen D,	
XX	Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gierzen D,	
PI	Patry S, Shi X, Suarez CJ,	
XX	Patry S, Shi X, Suarez CJ,	
DR	WPI: 2004-129368/30.	
XX	WPI: 2004-129368/30.	
DR	N-PSDB; ACN42164.	
XX	N-PSDB; ACN42164.	
PT	New diagnostic and therapeutic polynucleotides and polypeptides, useful	
XX	New diagnostic and therapeutic polynucleotides and polypeptides, useful	
PT	in diagnosing a condition, disease or disorder associated with human	
XX	in diagnosing a condition, disease or disorder associated with human	
PT	molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or	
XX	molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or	
PS	in gene mapping.	
XX	in gene mapping.	
PS	Claim 27; Page: 190pp; English.	
XX	Claim 27; Page: 190pp; English.	
CC	The invention relates to novel diagnostic and therapeutic polynucleotides	
XX	The invention relates to novel diagnostic and therapeutic polynucleotides	
CC	selected from one of the 722 sequences defined in the specification. A	
XX	selected from one of the 722 sequences defined in the specification. A	
CC	polynucleotide of the invention may have a use in gene therapy. The human	
XX	polynucleotide of the invention may have a use in gene therapy. The human	
CC	diagnostic and therapeutic polynucleotides (dittp) or polypeptides may be	
XX	diagnostic and therapeutic polynucleotides (dittp) or polypeptides may be	

used to diagnose a particular condition, disease or disorder associated with human molecules, e.g. cell proliferative disorders, endocrine disorders, neurological disorders, developmental disorder, endocrine infections caused by virus, bacteria, fungi or parasite. The dthp molecules may also be used in genetic mapping, in identifying individuals from minute biological samples, in detecting single nucleotide polymorphisms, as molecular weight markers, and for somatic or germline gene therapy. The present sequence represents a dthp protein of the invention. Note: The sequence data for this patent is not represented in the printed specification, but was obtained in electronic format directly from WIPO at www.wipo.int/pct/en/sequences/listing.htm

Sequence 598 AA;

Query Match 92.9%; Score 39; DB 8; Length 598;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 KXHPQEK 8
:|||||
Db 417 KTHPOYER 424

RESULT 2
ID AAM59212 standard; peptide; 8 AA.

AAW59212;

27-AUG-1998 (first entry)

Streptavidin tagged peptide ligand #2.

Streptavidin; ligand; binding affinity; mutant; isolation; purification; recover; immobilise.

Synthetic.

EP835934-A2.

15-APR-1998.

09-OCT-1997; 97EP-00117504.

10-OCT-1996; 96DE-01041876.

(BIOA-) INST BIOANALYTIK GMBH.

Skerra A, Voss S;

WPI; 1998-218868/20.

Streptavidin mutants with higher binding affinity for peptide ligands - have mutation in amino acid region 44-53, used to isolate, purify or determine fusion proteins including these ligands.

Claim 10; Page 11; 21pp; German.

AAW59211 and AAW59212 are ligands used in a method to assay binding affinity of streptavidin mutants. These mutants have a mutation within the amino acid (aa) region 44-53 of the wild-type protein show a higher binding affinity than the wild-type for peptide ligands that include the sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin mutants can be used to isolate, purify and determine proteins or to determine/recover substances that contain streptavidin-binding groups. Such compounds may also be used to immobilise fusions on microtitre plates, microbeads or sensor chips

Sequence 8 AA;

Query Match 88.1%; Score 37; DB 2; Length 8;

Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQEK 8
:|||||
Db 2 SHPOFER 8

RESULT 3
ID AAB35433 standard; peptide; 8 AA.

AAB35433;

23-MAY-2001 (first entry)

Epitope peptide #3.

Nascent protein detection; protein analysis; aminoacylated tRNA; BODIPY marker; disease diagnosis.

Unidentified.

WO200114578-A1.

01-MAR-2001.

23-AUG-2000; 2000WO-US023233.

25-AUG-1999; 99US-00382736.

25-AUG-1999; 99US-00382950.

(AMBE-) AMBERGEN INC.

Rothschild KJ, Gile S, Olejnik J;

WPI; 2001-168972/17.

Method for detecting nascent proteins by fluorescence comprises aminoacylating a tRNA molecule with a marker compound, useful for detecting mutations in proteins, e.g. cancer.

Disclosure; Page 47; 204pp; English.

The present invention describes a method of detecting nascent proteins involving aminoacylating a tRNA molecule with a 4',4'-difluoro-4-bora-3a,4a-diaza-8-indacene (BODIPY) marker leading to the production of a misaminoacylated tRNA. This enables the detection, isolation and analysis of nascent proteins using UV without the usual accompanying radioactivity problems. It may be used to detect mutations, for example in cancer, Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer

Sequence 8 AA;

Query Match 88.1%; Score 37; DB 4; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQEK 8
:|||||
Db 2 SHPOFER 8

RESULT 4

ID AAB35442 standard; peptide; 8 AA.

AAB35442;

23-MAY-2001 (first entry)

Nascent protein detection method related peptide #4.

```
KW Nascent protein detection; protein analysis; aminoacylated tRNA;
KW BODIPY marker; disease diagnosis.
XX Unidentified.
XX MO200114578-A1.
XX
XX PD 01-MAR-2001.
XX
XX PF 23-AUG-2000; 2000WO-US023223.
XX
XX PR 25-AUG-1999; 99US-00382736.
XX PR 25-AUG-1999; 99US-00382950.
XX
XX PA (AMBE-) AMBERGEN INC.
XX
XX PI Roehschild KJ, Gite S, Olejnik J;
XX
XX DR WPI, 2001-168972/17.
XX
XX PT Method for detecting nascent proteins by fluorescence comprises
XX PT misaminoacylating a tRNA molecule with a marker compound, useful for
XX PT detecting mutations in proteins, e.g. cancer.
XX
XX PS Example 22; Page 153; 204pp; English.
XX
XX The present invention describes a method of detecting nascent proteins
XX involving aminoacylating a tRNA molecule with a 4,4-difluoro-4-bora-3a,4a
XX -diaz-a-8-indacene (BODIPY) marker leading to the production of a
XX misaminoacylated tRNA. This enables the detection, isolation and analysis
XX of nascent proteins using UV without the usual accompanying radioactivity
XX problems. It may be used to detect mutations, for example in cancer,
XX Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer
XX
XX SQ Sequence 8 AA;

Query Match      88.1%; Score 37; DB 4; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFEK 8
       :|||||
DB      2 SHPQFEK 8

RESULT 5
AAB68616
ID AAB68616 standard; peptide; 8 AA.
XX
AC AAB68616;
XX
XX DT 27-APR-2001 (first entry)
XX
XX DE Strep-Tag II sequence.
XX
XX KM Protein-RNA fusion; Strep-Tag II.
XX
XX OS Unidentified.
XX
XX PN WO200107657-A1.
XX
XX PD 01-FEB-2001.
XX
XX PF 19-JUL-2000; 2000WO-US019653.
XX
XX PR 27-JUL-1999; 99US-0145834P.
XX
XX PA (PHYL-) PHYLLOS INC.
XX
XX PI Kurz M, Lohse P, Wagner R;
XX
XX DR WPI, 2001-182803/18.
XX
```

```
PT Affixing a peptide acceptor to an RNA molecule useful for producing
PT fusion proteins for isolating proteins or nucleic acids with desired
PT properties through attachment of a peptide acceptor to the 3' end of an
PT RNA molecule.
XX
XX PS Example 5; Page 22; 56pp; English.
XX
XX CC The present invention relates to a method for affixing a peptide acceptor
XX CC to an RNA molecule through the formation of a covalent bond, noncovalent
XX CC bond, or by chemical ligation. The method is useful for producing RNA-
XX CC protein fusions which can be used for the isolation of proteins or
XX CC nucleic acids with desired properties from large pools of partially or
XX CC completely random amino acid or nucleic acid sequences. The present
XX CC sequence is a Strep-Tag II used in the present invention
XX
XX SQ Sequence 8 AA;

Query Match      88.1%; Score 37; DB 4; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFEK 8
       :|||||
DB      2 SHPQFEK 8

RESULT 6
AA019065
ID AA019065 standard; peptide; 8 AA.
XX
XX AC AA019065;
XX
XX DT 14-NOV-2002 (first entry)
XX
XX DE Mutation detection method tag peptide SEQ ID NO: 34.
XX
XX KM Mutation detection; primer; mutant; tag; tumour suppressor gene;
XX KM protein production; cancer.
XX
XX OS Synthetic.
XX
XX PN WO20026675-A2.
XX
XX PD 29-AUG-2002.
XX
XX PF 15-FEB-2002; 2002WO-EP001651.
XX
XX PR 16-FEB-2001; 2001DE-01007317.
XX
XX PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
XX PI Kahmann S, Mueller O;
XX
XX DR WPI: 2002-674959/72.
XX DR N-PSDB; AAL49464.
XX
XX PT Detecting mutations in nucleic acid, useful for diagnosis and
XX PT characterization of tumors, by amplification, in vitro transcription and
XX PT translation, then protein detection.
XX
XX PS Disclosure; Fig 5; 62pp; German.
XX
XX CC The present invention relates to a method of detecting mutations in a
XX CC nucleic acid by amplifying the nucleic acid to produce a double-stranded
XX CC amplicon, in vitro transcription and translation of this amplicon, and
XX CC detection of the translated protein. The primers used for amplification
XX CC are designed to produce an amplicon that is translatable and allows
XX CC differentiation between translation products of wild-type and mutated
XX CC nucleic acids. The method is used to detect mutations in tumour
XX CC suppressor genes, for (early) diagnosis, monitoring and characterisation
XX CC of tumours (especially of bladder and intestines) and in the germ line
XX CC (using nucleic acids from embryos or blood cells). A new multi-tag vector
XX CC is used to detect or verify the reading frame of a nucleic acid cloned in
```

CC it, and to determine the suitability of detectable peptides for analysis
CC and/or purification of a recombinant protein, expressed from a sequence
CC cloned in the vector. The present sequence is a tag peptide which was
CC used in the invention
XX
SQ Sequence 8 AA;
Query Match 88.1%; Score 37; DB 5; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 XHPQFEK 8
:|||||
Db 2 SHPQFEK 8
RESULT 7
ABBS7464
ID ABB57464 standard; peptide; 8 AA.
XX
AC ABB57464;
XX
DT 18-MAR-2002 (first entry)
XX
DE STRAP tag II peptide.
XX
KW Immunomodulatory human MHC class II antigen-binding protein; HLA;
KW human leukocyte antigen; immune system; immunosuppression; antibody;
KW major histocompatibility complex; antidiabetic; antiarthritic;
KW neuroprotective; antiinflammatory; antipostarctic;
KW immunosuppressive; dermatological; antihypertensive; narcolepsy;
KW chymotrypsin; hepatocellular; immune response suppressor; insulin;
KW rheumatoid arthritis; juvenile arthritis; multiple sclerosis; insulin;
KW Grave's disease; insulin-dependent diabetes; Hashimoto's disease;
KW systemic lupus erythematosus; ankylosing spondylitis; myasthenia gravis;
KW transplant rejection; graft versus host disease; pemphigus vulgaris;
KW glomerulonephritis; thyroiditis; pancreatitis; primary biliary cirrhosis;
KW irritable bowel disease; Sjogren's syndrome.
XX
OS Synthetic.
XX
PN MO200187338-A1.
XX
PD 22-NOV-2001.
XX
PF 14-MAY-2001; 2001WO-US015626.
XX
PR 12-MAY-2000; 2000EP-00110063.
XX 06-OCT-2000; 2000US-0238762P.
XX
PA (GPCB-) GPC BIOTECH AG.
XX (MORP-) MORPHOSYS AG.
XX
PI Nagy Z, Tesar M, Thomassen-Wolf E;
XX
DR WPI; 2002-075289/10.
XX
PT Composition for suppressing immune response, treating diseases of immune
PT system, has polypeptide comprising antibody-based antigen-binding domain
XX of human composition, which binds antigen expressed on a cell surface.
XX
PS Disclosure; Page 31; 139pp; English.
XX
CC The present invention describes a composition (I), comprising a
CC polypeptide comprising an antibody-based antigen-binding domain of human
CC composition with binding specificity for an antigen expressed on the
CC surface of a cell, where treating cells expressing the antigen with the
CC polypeptides leads to suppression of an immune response, and the IC50 for
CC the suppression of immune response is 1 microm or less. (I) has
CC antineoplastic, antiarthritic, neuroprotective, antiinflammatory,
CC antidiabetic, antipostarctic, immunosuppressive, dermatological,
CC antihypertensive, hepatocellular, immunosuppressive, and hepatotropic activities, and
CC can be used as a suppressor of immune response. (I) is useful for

CC suppressing activation or proliferation of a cell of the immune system,
CC suppressing IL-2 secretion by a cell, the interaction of a cell of the
CC immune system with another cell, immunosuppression of a patient and for
CC killing a cell expressing an antigen, human leukocyte antigen (HLA)-DR on
CC the surface of the cell, where neither cytotoxic entities nor
CC immunological mechanisms are needed to cause or lead to the killing. (I)
CC (optionally linked to cytotoxic or immunogenic agent) is useful for
CC preparing a pharmaceutical preparation for the treatment of rheumatoid
CC arthritis, juvenile arthritis, multiple sclerosis, Grave's disease,
CC insulin-dependent diabetes, narcolepsy, psoriasis, systemic lupus
CC erythematosus, ankylosing spondylitis, transplant rejection, graft versus
CC host disease, Hashimoto's disease, myasthenia gravis, pemphigus vulgaris,
CC glomerulonephritis, thyroiditis, pancreatitis, insulin, primary biliary
CC cirrhosis, irritable bowel disease and Sjogren's syndrome in humans.
CC ABA92469 to ABA92474 and ABB57457 to ABB57590 represent sequence used in
CC the exemplification of the present invention
XX
SQ Sequence 8 AA;
Query Match 88.1%; Score 37; DB 5; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 XHPQFEK 8
:|||||
Db 2 SHPQFEK 8
RESULT 8
ABBS7486
ID ABB7486 standard; peptide; 8 AA.
XX
AC ABB7486;
XX
DT 22-JUN-2002 (first entry)
XX
DE AAV-helper plasmid related strep-peptide tag SEQ ID NO 17.
XX
KW Adeno-associated virus; AAV; glycoprotein; ITR; cap; rep; vector.
XX
OS Synthetic.
XX
PN MO200238782-A2.
XX
PD 16-MAY-2002.
XX
PF 13-NOV-2001; 2001WO-EP013125.
XX 13-NOV-2000; 2000DE-01056210.
XX
PA (ARIM-) ARIMEDES BIOTECHNOLOGY GMBH.
XX
PI Oberberger G, Hellmuth K, Wegener C;
XX
DR WPI; 2002-435853/46.
XX
PT Vector system for preparing recombinant adeno-associated viral particles,
PT used for high-level expression of heterologous therapeutic proteins in
XX eukaryotic cells.
XX
PS Disclosure; Page 17; 59pp; German.
XX
CC The invention relates to a viral vector system for preparing recombinant
CC adeno-associated virus (AAV) particles comprising; at least two plasmid
CC vectors (AB158983 and AB158984) that include the two inverted terminal
CC repeats (ITR) of AAV and additional sequences and plasmid vectors without
CC ITRs but containing the rep and cap genes of AAV required for replication
CC and packaging. The system is useful for producing recombinant AAV for
CC production of a wide range of therapeutic glycoproteins in eukaryotic
CC cells. The system provides efficient, large scale production of
CC heterologous proteins in mammalian cells, without requiring an adenovirus
CC helper. It is not toxic to host cells and does not cause lysis, so
CC produced proteins are highly pure. The present sequence is that of a


```

CC peptide tag, useful to the invention
XX
SQ Sequence 8 AA:
      Query Match      88.1%; Score 37; DB 5; Length 8;
      Best Local Similarity 85.7%; Pred. No. 1.8e+06;
      Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY      2 XHPQFEK 8
      :|||||
      Db          2 SHPQFEK 8

RESULT 9
ABG73584
ID ABG73584 standard; protein; 8 AA.
XX
AC ABG73584;
XX
DT 03-MAR-2003 (first entry)
XX
DE StreptagII affinity peptide.
XX
KW Non-protein L-amino acid; O-acetyl-L-serine; agrochemical production;
nucleophile; O-acetyl-L-serine sulphydrylase; pharmaceutical production.
XX
OS Synthetic.
XX
PN EP1247869-A1.
XX
PD 09-OCT-2002;
XX
PF 28-MAR-2002; 2002EP-00007262.
XX
PR 04-APR-2001; 2001DE-01016881.
XX
PR 03-MAY-2001; 2001DE-01021515.
XX
PA (CONE ) CONDORTIUM ELEKTROCHEM IND GMBH.
XX
PI Maier T, Gaebert C;
XX
DR WPI; 2003-077522/08.
XX
PT Production of non-protein L-amino acids useful for the manufacture of
pharmaceuticals and agrochemicals, comprises an enzyme catalyzed reaction
of O-acetyl-L-serine with a nucleophile.
XX
PS Example 1; Page 8; 20pp; German.
XX
CC This invention describes a novel method for the production of non-protein
L-amino acids by the reaction of O-acetyl-L-serine with a nucleophile in
the presence of a catalyst comprising O-acetyl-L-serine sulphydrylase at
pH 5-7.4. The method of the invention is useful for the manufacture of
pharmaceuticals and agrochemicals. In contrast to the process described
in DE 10046934, a high nucleophile concentration can be used which
includes toxic compounds. This sequence represents an affinity peptide
containing a StreptagII motif which is used in the construction of fusion
genes containing Escherichia coli cysK and cysM fragments
XX
SQ Sequence 8 AA;
      Query Match      88.1%; Score 37; DB 6; Length 8;
      Best Local Similarity 85.7%; Pred. No. 1.8e+06;
      Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY      2 XHPQFEK 8
      :|||||
      Db          2 SHPQFEK 8

RESULT 10
AAE37230
ID AAE37230 standard; peptide; 8 AA.

```

```

XX
AC AAE37230;
XX
DT 07-AUG-2003 (first entry)
XX
DE Streptag II epitope peptide.
XX
KW Gene expression; therapy; isolation; epitope.
XX
OS Synthetic.
XX
PN WO2003038049-A2.
XX
PD 08-MAY-2003.
XX
PF 29-OCT-2002; 2002WO-US034645.
XX
PR 29-OCT-2001; 2001US-0340689P.
XX
PA (RENO-) RENOVIS INC.
XX
PI Heintz N, Serafini TA, Shyjan AW;
XX
DR WPI; 2003-430512/40.
XX
PT Isolating cell-type specific mRNAs, useful in gene expression analysis or
quantification in a specific cell in a heterogeneous cell mixture, by
isolating mRNA bound to ribosomes or proteins that bind mRNA in a cell
type specific manner.
XX
PS Example 2; Page 121; 136pp; English.
XX
CC The invention relates to a method for isolating mRNA from a population of
cells. The method involves selectively isolating ribosomes or proteins
that bind mRNA in a cell type specific manner and then isolating the mRNA
bound to the ribosomes or proteins that bind mRNA. The method is useful
for facilitating the analysis and quantification of gene expression in a
selected cell type present within a heterogeneous cell mixture. The
method may also be used in diagnostics or therapies for human diseases.
CC The present sequence is Streptag II epitope peptide. This sequence is
used to illustrate the method of the invention
XX
SQ Sequence 8 AA;
      Query Match      88.1%; Score 37; DB 6; Length 8;
      Best Local Similarity 85.7%; Pred. No. 1.8e+06;
      Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY      2 XHPQFEK 8
      :|||||
      Db          2 SHPQFEK 8

RESULT 11
ABP60361
ID ABP60361 standard; peptide; 8 AA.
XX
AC ABP60361;
XX
DT 28-MAR-2003 (first entry)
XX
DE Streptavidin 11 tag peptide SEQ ID NO 2.
XX
KW Streptavidin; protein chip; microtitre plate; detection.
XX
OS Synthetic.
XX
PN DE10113776-A1.
XX
PD 02-OCT-2002.
XX
PF 21-MAR-2001; 2001DE-01013776.
XX

```

PR 21-MAR-2001; 2001DE-01013776.
XX (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX Schmidt T;
XX WPI; 2003-031166/03.
XX
XX New isolated peptide, useful as affinity purification tag for recombinant
PT protein, comprises at least two high-affinity streptavidin-binding
PT modules.
PS Example 1; Page 8; 18pp; German.
XX
XX The invention relates to an isolated peptide (I) comprising at least two
CC individual modules separated by 0-50 amino acids, with each containing at
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
CC streptavidin binding modules, are useful as affinity handles for
CC purification of recombinant fusion proteins (FP), also for detecting FP,
CC e.g. on protein chips or microtitre plates. The modules in (I) bind
CC strongly to streptavidin, with a co-operative effect that provides
CC (I) does not interfere with the function of attached proteins (II) (so it
CC may not be essential to remove it); facilitates detection and has easily
CC controllable binding properties. (I) is particularly used for purifying
CC FP from dilute solution in batch formats (which use simpler apparatus
CC than column methods and result in lower loss of FP). The present sequence
CC is that of a streptavidin tag peptide disclosed with the invention
SQ Sequence 8 AA;
OY Query Match 88.1%; Score 37; DB 6; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
DB 2 XHPQPEK 8
:|||||
2 SHPQPEK 8
RESULT 12
ABP60368
ID ABP60368 standard; peptide; 8 AA.
XX
XX ABP60368;
AC
XX 28-MAR-2003 (first entry)
DT
XX Streptavidin binding peptide SEQ ID NO 9.
DE
XX Streptavidin; protein chip; microtitre plate; detection.
KW
XX Synthetic.
OS
XX DE10113776-A1.
PN
XX 02-OCT-2002.
PD
XX 21-MAR-2001; 2001DE-01013776.
PF
XX 21-MAR-2001; 2001DE-01013776.
PR
XX (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
PA
XX Schmidt T;
PI
XX WPI; 2003-031166/03.
DR
XX New isolated peptide, useful as affinity purification tag for recombinant
PT protein, comprises at least two high-affinity streptavidin-binding
PT modules.
PS Claim 5; Page 16; 18pp; German.

XX The invention relates to an isolated peptide (I) comprising at least two
CC individual modules separated by 0-50 amino acids, with each containing at
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
CC streptavidin binding modules, are useful as affinity handles for
CC purification of recombinant fusion proteins (FP), also for detecting FP,
CC e.g. on protein chips or microtitre plates. The modules in (I) bind
CC strongly to streptavidin, with a co-operative effect that provides
CC stronger binding than a single tag, but are displaced by a competitor.
CC (I) does not interfere with the function of attached proteins (II) (so it
CC may not be essential to remove it); facilitates detection and has easily
CC controllable binding properties. (I) is particularly used for purifying
CC FP from dilute solution in batch formats (which use simpler apparatus
CC than column methods and result in lower loss of FP). The present sequence
CC is that of a streptavidin binding peptide disclosed with the invention
SQ Sequence 9 AA;
OY Query Match 88.1%; Score 37; DB 6; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
DB 2 XHPQPEK 8
:|||||
2 SHPQPEK 8
RESULT 13
ADA09808
ID ADA09808 standard; peptide; 8 AA.
XX
XX ADA09808;
AC
XX 06-NOV-2003 (first entry)
DT
XX StepTag epitope useful as an affinity marker.
DE
XX Non-radioactive marker; nascent protein detection method;
KW cellular translation system; cell-free translation system;
KW dipyrrometheneboron difluoride dye;
KW 4,4'-difluoro-4-bora-3a,4a-diaza-8-indacene dye; vaccine; drug;
KW human disease screening; human disorder; protein separation;
KW affinity marker; StepTag epitope.
XX
XX Synthetic.
OS
XX US2003092031-A1.
PN
XX 15-MAY-2003.
PD
XX 18-JUN-2002; 2002US-00174368.
PF
XX 25-AUG-1999; 99US-00382736.
PR 23-AUG-2000; 2000WO-US023233.
PR 21-JUN-2002; 2002US-00049332.
XX
XX (AMBE-) AMBERGEN INC.
PA
XX Rothechild KJ, Gite S, Olejnik J;
PI
XX WPI; 2003-576764/54.
DR
XX Detecting, analyzing or isolating nascent proteins comprises introducing
PT a modified nucleic acid template into a cellular or cell-free translation
PT system to generate a nascent protein having at least an N-terminal
PT marker.
PS Disclosure; Page 15; 76pp; English.
XX The present invention relates to non-radioactive markers used in the
CC detection and analysis of nascent proteins translated in cellular or cell
CC free translation systems. The preferred non-radioactive markers are
CC dipyrrometheneboron difluoride (4,4'-difluoro-4-bora-3a,4a-diaza-8-

CC Indacene) dyes. The detection method of the invention is a gel-free
 CC method that comprises introducing a modified nucleic acid template into a
 CC translation system under conditions such that a nascent protein is
 CC generated, the protein comprising at least an N-terminal marker. The
 CC method is useful in detecting, analysing and isolating nascent proteins
 CC produced in a cell-free or cellular translation system without the use of
 CC radioactive amino acids or other radioactive labels. Compositions
 CC comprising nascent proteins translated in the presence of markers may be
 CC used as vaccines or as drugs for humans and other animals. The method and
 CC a kit containing reagents for the detection of nascent proteins may be
 CC used as a rapid means to screen humans or other animals for the presence
 CC of certain diseases or disorders. The present sequence represents an
 CC epitope that can be used as an affinity marker for protein separation.

CC Sequence 8 AA;

Query Match 88.1%; Score 37; DB 7; Length 8;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQEK 8
 :|||||
 DB 2 SHPQEK 8

RESULT 14

ID ADB84588 standard; peptide; 8 AA.

AC ADB84588;

DT 04-DEC-2003 (first entry)

DE Streptavidin conserved peptide #2.

KW cell-free transcription system; cell-free translation system;
 protein synthesis; matrix; streptavidin.

OS Escherichia coli.

PN DE10137792-A1.

PD 27-FEB-2003.

PF 06-AUG-2001; 2001DE-01037792.

PR 06-AUG-2001; 2001DE-01037792.

PA (ERDM/) ERDMANN V.

PI Erdmann VA, Lamla T, Stiege W;

DR WPI; 2003-343999/33.

PT Expressing genes in cell-free system, useful for preparation of proteins,
 PT comprising that the protein formed is removed from solution by binding to
 PT a matrix.

PS Claim 13; Col 8; 8pp; German.

XX This invention describes a novel method of expressing genes in a cell-
 CC free transcription and translation system which comprises using a
 CC reaction solution containing all necessary components of the
 CC transcription/translation system, amino acids, nucleotides and
 CC metabolites that supply energy and that are needed for synthesis. The
 CC proteins formed are immobilised on a matrix. The method allows simple
 CC recovery of proteins without a separate isolation step and the amount of
 CC proteins produced can be determined before a reaction is complete.
 CC Continuous removal of proteins prevents it interfering with the
 CC expression process, making possible synthesis of proteins that interact
 CC adversely with the process, so normally produced only in very low yields.
 CC The use of a matrix concentrates the proteins produced and is applicable
 CC to proteins of any size. This sequence represents a highly conserved

CC peptide from E. coli streptavidin which is used to illustrate the method
 CC of the invention.

CC Sequence 8 AA;

Query Match 88.1%; Score 37; DB 7; Length 8;
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQEK 8
 :|||||
 DB 2 SHPQEK 8

RESULT 15

ID ADB85500 standard; peptide; 8 AA.

AC ADB85500;

DT 04-DEC-2003 (first entry)

DE Streptavidin tag peptide related to human aggrecanase.

KW aggrecanase; aggrecan; articular cartilage; proteoglycan;
 KW cartilage compressibility; cartilage elasticity; arthritic disease;
 KW osteoarthritis; cartilage degradation; inflammatory joint disease;
 KW aggrecanase-1; ADAMTS-4; aggrecanase-2; ADAMTS-5; thrombospondin domain;
 KW TSP domain; osteopathic; antiarthritic; cytostatic; antiinflammatory;
 KW antirheumatic; ophthalmological; thrombolytic; vasotropic; antimicrobial;
 KW respiratory-gen; nocotropic; neuroprotective; antiparkinsonian;
 KW immunosuppressive; aggrecanase inhibition; cancer; rheumatoid arthritis;
 KW septic arthritis; corneal ulceration; coronary thrombosis;
 KW Crohn's disease; emphysema; Alzheimer's disease; Parkinson's disease;
 KW multiple sclerosis; aortic aneurysm; streptavidin tag.

OS Synthetic.

PN WO2003066822-A2.

PD 14-AUG-2003.

PF 05-FEB-2003; 2003WO-US003554.

PR 05-FEB-2002; 2002US-0354592P.

PA (AMHP) WYETH.

PI Georgiadis K, Crawford TK, Tomkinson KM, Morris EA, Racie L;

DR WPI; 2003-731495/69.

DR N-PSDB; ADB85498.

PT New biologically-active aggrecanase protein having a deletion of all, or
 PT a portion of a TSP domain, useful for treating osteoarthritis, cancer,
 PT Parkinson's disease, coronary thrombosis, Alzheimer's disease and
 PT multiple sclerosis.

PS Example 1; Fig 1a; 11pp; English.

XX This invention relates to novel truncated human aggrecanase proteins and
 CC nucleotide sequences. Aggrecan is a major extracellular component of
 CC articular cartilage. It is a proteoglycan responsible for providing
 CC cartilage with its mechanical properties of compressibility and
 CC elasticity. The loss of aggrecan has been implicated in the degradation
 CC of articular cartilage in arthritic diseases such as osteoarthritis.
 CC Aggrecanase is responsible for the cleavage of aggrecan, thereby having a
 CC role in cartilage degradation associated with osteoarthritis and
 CC inflammatory joint disease. The proteins of the current invention are
 CC truncated aggrecanase-1 (ADAMTS-4) and aggrecanase-2 (ADAMTS-5) enzymes
 CC which have at least one thrombospondin (TSP) domain deleted. These are
 CC biologically active and have greater stability and higher expression than
 CC their full-length counterparts. The proteins of the invention may be of

CC use in the development of compounds with osteopathic, antiarthritic,
CC cytosolic, antiinflammatory, antirheumatic, ophthalmological,
CC thrombolytic, vasotropic, antimicrobial, respiratory-gen, nootropic,
CC neuroprotective, antiparkinsonian or immunosuppressive activities through
CC aggregase inhibition. The proteins of the invention may therefore be
CC useful for the manufacture of compositions for the treatment of
CC aggregase-associated conditions, such as osteoarthritis, cancer,
CC inflammatory joint disease, rheumatoid arthritis, septic arthritis,
CC corneal ulceration, coronary thrombosis, Crohn's disease, emphysema,
CC Alzheimer's disease, Parkinson's disease, multiple sclerosis and aortic
CC aneurysm. The present sequence is that of a streptavidin tag peptide
CC linker which was used during the exemplification of the invention.

XX
SQ Sequence 8 AA;

Query Match 88.1%; Score 37; DB 7; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFEK 8
:|||||
Db 2 SHPQFEK 8

Search completed: March 2, 2005, 13:02:44
Job time : 36.0325 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 24.878 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ5
Perfect score: 42
Sequence: 1 kmhgafek 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	90.5	36	14	US-10-026-578B-11
2	37	88.1	8	9	US-09-809-517A-9
3	37	88.1	8	9	US-09-873-145-7
4	37	88.1	8	12	US-09-813-197-8
5	37	88.1	8	13	US-10-104-218-5
6	37	88.1	8	13	US-10-208-357-9
7	37	88.1	8	14	US-10-001-934-8
8	37	88.1	8	14	US-10-026-578B-2
9	37	88.1	8	14	US-10-026-578B-9
10	37	88.1	8	14	US-10-174-368A-7
11	37	88.1	8	14	US-10-264-127-8
12	37	88.1	8	15	US-10-339-712-8
13	37	88.1	8	15	US-10-339-712-67

14	37	88.1	8	15	US-10-275-046-4	Sequence 4, Appli
15	37	88.1	8	15	US-10-425-000-76	Sequence 76, Appli
16	37	88.1	8	15	US-10-424-959-25	Sequence 25, Appli
17	37	88.1	8	15	US-10-358-283-23	Sequence 23, Appli
18	37	88.1	8	16	US-10-628-432-41	Sequence 41, Appli
19	37	88.1	8	17	US-10-494-248-17	Sequence 17, Appli
20	37	88.1	8	17	US-10-634-645-11	Sequence 11, Appli
21	37	88.1	8	17	US-10-719-523-8	Sequence 8, Appli
22	37	88.1	9	9	US-09-983-067-3	Sequence 3, Appli
23	37	88.1	10	9	US-09-809-517A-6	Sequence 6, Appli
24	37	88.1	10	15	US-10-147-211A-20	Sequence 20, Appli
25	37	88.1	11	15	US-10-354-983-29	Sequence 29, Appli
26	37	88.1	11	16	US-10-628-432-25	Sequence 25, Appli
27	37	88.1	21	9	US-09-809-517A-33	Sequence 30, Appli
28	37	88.1	22	9	US-09-809-517A-30	Sequence 31, Appli
29	37	88.1	24	9	US-09-809-517A-31	Sequence 33, Appli
30	37	88.1	24	14	US-10-026-578B-3	Sequence 3, Appli
31	37	88.1	24	14	US-10-026-578B-4	Sequence 4, Appli
32	37	88.1	25	9	US-09-809-517A-34	Sequence 34, Appli
33	37	88.1	36	14	US-10-026-578B-10	Sequence 10, Appli
34	37	88.1	117	10	US-09-977-137A-4	Sequence 4, Appli
35	37	88.1	117	10	US-09-977-137A-5	Sequence 5, Appli
36	37	88.1	117	10	US-09-977-137A-7	Sequence 7, Appli
37	37	88.1	117	10	US-09-977-137A-8	Sequence 8, Appli
38	37	88.1	117	10	US-09-977-137A-9	Sequence 9, Appli
39	37	88.1	117	10	US-09-977-137A-10	Sequence 10, Appli
40	37	88.1	117	10	US-09-977-137A-11	Sequence 11, Appli
41	37	88.1	117	10	US-09-977-137A-12	Sequence 12, Appli
42	37	88.1	118	10	US-09-977-137A-6	Sequence 6, Appli
43	37	88.1	245	17	US-10-887-228A-1	Sequence 9, Appli
44	37	88.1	246	17	US-10-887-228A-9	Sequence 9, Appli
45	37	88.1	252	17	US-10-887-228A-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-10-026-578B-11
; Sequence 11, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: Schmitt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(28)
; OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more
; OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
; US-10-026-578B-11

Query Match 90.5%; Score 38; DB 14; Length 36;
Best Local Similarity 75.0%; Pred. No. 12;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 KXHPQFEK 8
Db 29 QSHPOFEK 36

RESULT 2
US-09-809-517A-9
; Sequence 9, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-9

Query Match 88.1%; Score 37; DB 9; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8
Db 2 SHPOFEK 8

RESULT 3
US-09-973-145-7
; Sequence 7, Application US/09973145
; Patent No. US20020132248A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
; FILE REFERENCE: AMBR-06819
; CURRENT APPLICATION NUMBER: US/09/973,145
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 09/382,950
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; NAME/KEY: misc feature
; OTHER INFORMATION: Synthetic
US-09-973-145-7

Query Match 88.1%; Score 37; DB 9; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8

Db 2 SHPOFEK 8

RESULT 4
US-09-813-197-8
; Sequence 8, Application US/09813197
; Publication No. US20050009013A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot
; FILE REFERENCE: AMBR-03951
; CURRENT APPLICATION NUMBER: US/09/813,197
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-813-197-8

Query Match 88.1%; Score 37; DB 12; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8
Db 2 SHPOFEK 8

RESULT 5
US-10-104-218-5
; Sequence 5, Application US/10104218
; Publication No. US20020177196A1
; GENERAL INFORMATION:
; APPLICANT: MAIER, Thomas
; APPLICANT: GAEBERT, Carsten
; TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
; FILE REFERENCE: MAIER, T. ET AL.-2
; CURRENT APPLICATION NUMBER: US/10/104,218
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 16 881.0
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 21 515.0
; PRIOR FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(8)
; OTHER INFORMATION: StreptagII affinity peptide for protein purification
US-10-104-218-5

Query Match 88.1%; Score 37; DB 13; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8
Db 2 SHPOFEK 8

RESULT 6
US-10-208-357-9

```
; Sequence 9, Application US/10208357
; Publication No. US20020182687A1
; GENERAL INFORMATION:
; APPLICANT: Kurtz, Markus
; APPLICANT: Lohse, Peter
; APPLICANT: Wagner, Richard
; TITLE OF INVENTION: Peptide Acceptor Ligation Methods
; FILE REFERENCE: 50036/031002
; CURRENT APPLICATION NUMBER: US/10/208,357
; CURRENT FILING DATE: 2002-07-30
; PRIOR APPLICATION NUMBER: US/09/619,103
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/145,834
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: designed sequence to act as an identifying tag
US-10-208-357-9

Query Match      88.1%; Score 37; DB 13; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQEK 8
       :|||||
Db      2 SHPQEK 8

RESULT 7
US-10-001-934-8
; Sequence 8, Application US/10001934
; Publication No. US20030032782A1
; GENERAL INFORMATION:
; APPLICANT: NAGY, ZOLTAN
; TITLE OF INVENTION: HUMAN PEPTIDES/PROTEINS CAUSING OR LEADING TO THE
; TITLE OF INVENTION: KILLING OF CELLS INCLUDING LYMPHOID TUMOR CELLS
; FILE REFERENCE: GPCG-P01-003
; CURRENT APPLICATION NUMBER: US/10/001,934
; CURRENT FILING DATE: 2001-11-15
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-001-934-8

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQEK 8
       :|||||
Db      2 SHPQEK 8

RESULT 8
US-10-026-578B-2
; Sequence 2, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
```

```
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: misc.feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-2

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQEK 8
       :|||||
Db      2 SHPQEK 8

RESULT 9
US-10-026-578B-9
; Sequence 9, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: Schmidt, Thomas
; APPLICANT: IBA (GmbH)
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: misc.feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-9

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQEK 8
       :|||||
Db      2 SHPQEK 8

RESULT 10
US-10-174-368A-7
; Sequence 7, Application US/10174368A
; Publication No. US20030092031A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gite, Sadehand
; APPLICANT: Olejnik, Jerzy
```

```

; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Pro
; FILE REFERENCE: AMBER-07145
; CURRENT APPLICATION NUMBER: US/10/174,368A
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: 10/049,332
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/23233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; NAME/KEY: misc feature
; OTHER INFORMATION: Synthetic
US-10-174-368A-7

```

```

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 XHPQFEK 8
   :|||||
Db 2 SHPQFEK 8

```

```

RESULT 11
US-10-264-127-8
; Sequence 8, Application US/10264127
; Publication No. US20030190643A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Pro
; FILE REFERENCE: AMBER-03951
; CURRENT APPLICATION NUMBER: US/10/264,127
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: US/09/382,736B
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-264-127-8

```

```

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 XHPQFEK 8
   :|||||
Db 2 SHPQFEK 8

```

```

RESULT 12
US-10-339-712-8
; Sequence 8, Application US/10339712
; Publication No. US20040014071A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy

```

```

; APPLICANT: Lim, Mark
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent
; FILE REFERENCE: AMBER-07199
; CURRENT APPLICATION NUMBER: US/10/339,712
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: 10/049,322
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/23233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-339-712-8

```

```

Query Match      88.1%; Score 37; DB 15; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 XHPQFEK 8
   :|||||
Db 2 SHPQFEK 8

```

```

RESULT 13
US-10-339-712-67
; Sequence 67, Application US/10339712
; Publication No. US20040014071A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent
; FILE REFERENCE: AMBER-07199
; CURRENT APPLICATION NUMBER: US/10/339,712
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: 10/049,322
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/23233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 67
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-339-712-67

```

```

Query Match      88.1%; Score 37; DB 15; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 XHPQFEK 8
   :|||||
Db 2 SHPQFEK 8

```

```

RESULT 14
US-10-275-046-4
; Sequence 4, Application US/10275046

```



```

; Publication No. US20040019187A1
; GENERAL INFORMATION:
; APPLICANT: Nagy et al.
; TITLE OF INVENTION: IMMUNOMODULATORY HUMAN MHC CLASS II ANTIGEN-BINDING POLYPEPTIDES
; FILE REFERENCE: GPCG-P01-260
; CURRENT APPLICATION NUMBER: US/10/275,046
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 00110063.5
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 4
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: STREP tag II
US-10-275-046-4

```

```

Query Match      88.1%; Score 37; DB 15; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      2 XHPQPEK 8
       :|||||
Db      2 SHPQPEK 8

```

```

RESULT 15
US-10-425-000-76
; Sequence 76, Application US/10425000
; Publication No: US20040052777A1
; GENERAL INFORMATION:
; APPLICANT: Neqbit, Mark
; APPLICANT: Cameron, Beatrice
; APPLICANT: Blanche, Francis
; TITLE OF INVENTION: Kringles Polypeptides and Methods for Using Them to Inhibit
; TITLE OF INVENTION: Angiogenesis
; FILE REFERENCE: ST01027-B
; CURRENT APPLICATION NUMBER: US/10/425,000
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: 10/233,675
; PRIOR FILING DATE: 2002-09-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 76
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Purification tag
US-10-425-000-76

```

```

Query Match      88.1%; Score 37; DB 15; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      2 XHPQPEK 8
       :|||||
Db      2 SHPQPEK 8

```

```

Search completed: March 2, 2005, 14:18:50
Job time : 24.878 sec

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 9.10569 Seconds
(without alignments)
65.585 Million cell updates/sec

Title: SEQ5

Perfect score: 42

Sequence: 1 kxhpqfek 8

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	88.1	8	US-08-948-097-2	Sequence 2, Appl1
2	37	88.1	8	US-09-382-950-7	Sequence 7, Appl1
3	37	88.1	8	US-09-382-736B-8	Sequence 8, Appl1
4	37	88.1	8	US-09-619-103-9	Sequence 9, Appl1
5	37	88.1	8	US-10-104-218-5	Sequence 5, Appl1
6	37	88.1	8	US-09-809-517A-9	Sequence 9, Appl1
7	37	88.1	10	US-09-809-517A-6	Sequence 9, Appl1
8	37	88.1	21	US-09-809-517A-30	Sequence 6, Appl1
9	37	88.1	22	US-09-809-517A-33	Sequence 30, Appl1
10	37	88.1	24	US-09-809-517A-31	Sequence 31, Appl1
11	37	88.1	25	US-09-809-517A-34	Sequence 34, Appl1
12	37	88.1	117	US-09-977-137A-4	Sequence 4, Appl1
13	37	88.1	117	US-09-977-137A-5	Sequence 5, Appl1
14	37	88.1	117	US-09-977-137A-7	Sequence 7, Appl1
15	37	88.1	117	US-09-977-137A-8	Sequence 8, Appl1
16	37	88.1	117	US-09-977-137A-9	Sequence 9, Appl1
17	37	88.1	117	US-09-977-137A-10	Sequence 10, Appl1
18	37	88.1	117	US-09-977-137A-11	Sequence 11, Appl1
19	37	88.1	117	US-09-977-137A-12	Sequence 12, Appl1
20	37	88.1	118	US-09-977-137A-6	Sequence 6, Appl1
21	36	88.7	482	US-09-107-532A-4868	Sequence 4868, Ap
22	35	83.3	250	US-09-803-286A-7	Sequence 7, Appl1
23	35	83.3	312	US-09-538-092-742	Sequence 742, App
24	35	83.3	1220	US-08-930-966A-2	Sequence 2, Appl1
25	34	81.0	180	US-09-134-000C-4700	Sequence 4700, Ap
26	34	81.0	207	US-09-248-796A-22051	Sequence 22051, A
27	34	81.0	213	US-09-107-532A-3797	Sequence 3797, Ap

28	34	81.0	234	4	US-09-270-767-32857	Sequence 32857, A
29	34	81.0	234	4	US-09-270-767-48074	Sequence 48074, A
30	34	81.0	410	4	US-09-248-796A-18368	Sequence 18368, A
31	34	81.0	428	3	US-09-052-778-12	Sequence 12, Appl1
32	34	81.0	608	4	US-09-270-767-32937	Sequence 32937, A
33	34	81.0	608	4	US-09-270-767-48154	Sequence 48154, A
34	34	81.0	712	4	US-09-248-796A-18407	Sequence 18407, A
35	34	81.0	979	2	US-08-308-881-6	Sequence 6, Appl1
36	34	81.0	979	2	US-09-058-263-6	Sequence 6, Appl1
37	34	81.0	979	2	US-09-058-099-6	Sequence 6, Appl1
38	34	81.0	979	3	US-09-058-264-6	Sequence 6, Appl1
39	34	81.0	979	4	US-09-455-962-6	Sequence 6, Appl1
40	34	81.0	979	5	PCR-US95-06530-6	Sequence 6, Appl1
41	33	78.6	114	4	US-09-270-767-42719	Sequence 42719, A
42	33	78.6	462	4	US-09-248-796A-18226	Sequence 18226, A
43	33	78.6	741	4	US-09-489-039A-8111	Sequence 8111, Ap
44	33	78.6	1040	2	US-08-254-989-2	Sequence 2, Appl1
45	33	78.6	1043	4	US-09-538-092-935	Sequence 935, Appl1

ALIGNMENTS

```
RESULT 1
US-08-948-097-2
; Sequence 2, Application US/08948097C
; Patent No. 6103493
; GENERAL INFORMATION:
; APPLICANT: Skerra, Arne
; APPLICANT: Voss, Selma
; TITLE OF INVENTION: Streptavidin Mutelins
; FILE REFERENCE: HUBR 119
; CURRENT FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: DE 196 41 876.3
; EARLIER FILING DATE: 1996-10-10
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: BINDING
; OTHER INFORMATION: Binding ligand for streptavidin
US-08-948-097-2

Query Match      88.1%; Score 37; DB 3; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFEK 8
      :|:|:|:|
DB      2 SHPOFEK 8

RESULT 2
US-09-382-950-7
; Sequence 7, Application US/09382950
; Patent No. 6303337
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
; FILE REFERENCE: AMBER-03879
; CURRENT APPLICATION NUMBER: US/09/382,950
; CURRENT FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent version 3.0
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial/Unknown
```

```
FEATURE:
NAME/KEY: misc feature
LOCATION: ( )
OTHER INFORMATION: Synthetic
US-09-382-950-7
```

```
Query Match      88.1%; Score 37; DB 3; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQFEK 8
Db      2 SHPQFEK 8
```

```
RESULT 3
US-09-382-736B-8
Sequence 8, Application US/09382736B
Patent No. 6306628
GENERAL INFORMATION:
```

```
APPLICANT: Rothschild, Kenneth
APPLICANT: Gite, Sadanand
```

```
TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Pro
FILE REFERENCE: AMBER-03951
```

```
CURRENT APPLICATION NUMBER: US/09/382,736B
CURRENT FILING DATE: 1999-08-25
```

```
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.0
```

```
SEQ ID NO 8
```

```
LENGTH: 8
```

```
TYPE: PRT
```

```
ORGANISM: Artificial Sequence
```

```
FEATURE:
```

```
OTHER INFORMATION: Synthetic
```

```
US-09-382-736B-8
Query Match      88.1%; Score 37; DB 3; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQFEK 8
Db      2 SHPQFEK 8
```

```
RESULT 4
US-09-619-103-9
```

```
Sequence 9, Application US/09619103
Patent No. 6429300
GENERAL INFORMATION:
```

```
APPLICANT: Kurz, Markus
APPLICANT: Lohse, Peter
```

```
TITLE OF INVENTION: Peptide Acceptor Ligation Methods
FILE REFERENCE: 50036/031002
```

```
CURRENT APPLICATION NUMBER: US/09/619,103
CURRENT FILING DATE: 2000-07-19
```

```
PRIOR APPLICATION NUMBER: 60/145,834
PRIOR FILING DATE: 1999-07-27
```

```
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 4.0
```

```
SEQ ID NO 9
```

```
LENGTH: 8
```

```
TYPE: PRT
```

```
ORGANISM: Artificial Sequence
```

```
FEATURE:
```

```
OTHER INFORMATION: designed sequence to act as an identifying tag
```

```
US-09-619-103-9
```

```
Query Match      88.1%; Score 37; DB 4; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQFEK 8
Db      2 SHPQFEK 8
```

```
RESULT 5
US-10-104-218-5
Sequence 5, Application US/10104218
Patent No. 6579705
GENERAL INFORMATION:
```

```
APPLICANT: MAIER, Thomas
APPLICANT: GABERT, Carsten
```

```
TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
FILE REFERENCE: MAIER, T. ET AL.-2
```

```
CURRENT APPLICATION NUMBER: US/10/104,218
CURRENT FILING DATE: 2002-03-22
```

```
PRIOR APPLICATION NUMBER: German No. 6579705 101 16 881.0
PRIOR FILING DATE: 2001-04-04
```

```
PRIOR APPLICATION NUMBER: German No. 6579705 101 21 515.0
PRIOR FILING DATE: 2001-05-03
```

```
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.0
```

```
SEQ ID NO 5
```

```
LENGTH: 8
```

```
TYPE: PRT
```

```
ORGANISM: Artificial Sequence
```

```
FEATURE:
```

```
NAME/KEY: PEPTIDE
```

```
LOCATION: (1)..(8)
```

```
OTHER INFORMATION: StreptagII affinity peptide for protein purification
```

```
US-10-104-218-5
Query Match      88.1%; Score 37; DB 4; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQFEK 8
Db      2 SHPQFEK 8
```

```
RESULT 6
US-09-809-517A-9
Sequence 9, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
```

```
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
```

```
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
```

```
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
```

```
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
```

```
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
```

```
SEQ ID NO 9
```

```
LENGTH: 8
```

```
TYPE: PRT
```

```
ORGANISM: artificial sequence
```

```
FEATURE:
```

```
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
```

```
US-09-809-517A-9
```

```
Query Match      88.1%; Score 37; DB 4; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQFEK 8
Db      2 SHPQFEK 8
```

Db 2 SHPOPEK 8

RESULT 7
US-09-809-517A-6
; Sequence 6, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 6
; LENGTH: 10
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-6

Query Match 88.1%; Score 37; DB 4; Length 10;
Best Local Similarity 85.7%; Pred. No. 0.54;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOPEK 8
:|||||
Db 4 SHPOPEK 10

RESULT 8
US-09-809-517A-30
; Sequence 30, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 21
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30

Query Match 88.1%; Score 37; DB 4; Length 21;
Best Local Similarity 85.7%; Pred. No. 1.2;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOPEK 8
:|||||
Db 15 SHPOPEK 21

RESULT 9
US-09-809-517A-33

; Sequence 33, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

Query Match 88.1%; Score 37; DB 4; Length 22;
Best Local Similarity 85.7%; Pred. No. 1.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOPEK 8
:|||||
Db 16 SHPOPEK 22

RESULT 10
US-09-809-517A-31
; Sequence 31, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
; LENGTH: 24
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31

Query Match 88.1%; Score 37; DB 4; Length 24;
Best Local Similarity 85.7%; Pred. No. 1.4;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOPEK 8
:|||||
Db 18 SHPOPEK 24

RESULT 11
US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

```

; TITLE OF INVENTION: particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match      88.1%; Score 37; DB 4; Length 25;
Best Local Similarity 85.7%; Pred. No. 1.5;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQPEK 8
Db      19 SHPQPEK 25

RESULT 12
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

Query Match      88.1%; Score 37; DB 4; Length 117;
Best Local Similarity 85.7%; Pred. No. 8;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQPEK 8
Db      111 SHPQPEK 117

RESULT 13
US-09-977-137A-5
; Sequence 5, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
```

```

; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5

Query Match      88.1%; Score 37; DB 4; Length 117;
Best Local Similarity 85.7%; Pred. No. 8;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQPEK 8
Db      111 SHPQPEK 117

RESULT 14
US-09-977-137A-7
; Sequence 7, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-7

Query Match      88.1%; Score 37; DB 4; Length 117;
Best Local Similarity 85.7%; Pred. No. 8;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQPEK 8
Db      111 SHPQPEK 117

RESULT 15
US-09-977-137A-8
; Sequence 8, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

```
! FEATURE:
! OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8

Query Match      88.1%; Score 37; DB 4; Length 117;
Best Local Similarity 85.7%; Pred. No. 8;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQPEK 8
      :|||||
Db      111 SHPQPEK 117
```

Search completed: March 2, 2005, 12:25:33
Job time : 9.10569 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OW protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 6.11382 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ6

Perfect score: 42

Sequence: 1 kxhpqfer 8

Scoring table:

BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	90.5	338	2	T30538
2	38	90.5	420	2	S71199
3	37	88.1	142	2	B82223
4	37	88.1	189	2	T36642
5	37	88.1	627	2	T00484
6	37	88.1	918	2	A55277
7	36	85.7	360	2	B71073
8	35	83.3	215	2	AG3392
9	35	83.3	397	2	S33312
10	35	83.3	413	2	T40427
11	34	81.0	149	2	T23939
12	34	81.0	249	2	S74762
13	34	81.0	258	2	E95998
14	34	81.0	300	2	G75436
15	34	81.0	313	2	H69297
16	34	81.0	331	2	T22648
17	34	81.0	336	2	T09133
18	34	81.0	339	2	B72402
19	34	81.0	354	2	S52579
20	34	81.0	355	2	T24938
21	34	81.0	360	2	T18140
22	34	81.0	362	2	H75131
23	34	81.0	376	1	HHBCDJ
24	34	81.0	376	2	G90630
25	34	81.0	376	2	G85481
26	34	81.0	377	2	F84947
27	34	81.0	377	2	JC5609
28	34	81.0	379	2	AF0503
29	34	81.0	379	2	AB0058

30	34	81.0	393	2	A55863	integrase - Strept
31	34	81.0	407	2	T39658	probable mitochond
32	34	81.0	413	2	S35581	dna protein homol
33	34	81.0	417	2	UQ2142	chaperone ANU1 pro
34	34	81.0	418	2	S42031	LDJ2 protein - lee
35	34	81.0	419	2	T07371	dna protein homol
36	34	81.0	420	2	T49127	dna protein homol
37	34	81.0	508	2	C82138	conserved hypochet
38	34	81.0	621	2	A71516	hypothetical prote
39	34	81.0	917	1	A31869	hexokinase (EC 2.7
40	34	81.0	918	1	A35244	hexokinase (EC 2.7
41	34	81.0	918	2	C59226	hexokinase (EC 2.7
42	34	81.0	1166	2	T15628	hypothetical prote
43	34	81.0	1220	2	T06403	resistance complex
44	33	78.6	101	2	T10856	carboxypeptidase C
45	33	78.6	115	2	T13519	hypothetical prote

ALIGNMENTS

RESULT 1

T30538

heat shock protein homolog dnaJ - Trypanosoma cruzi

C/Species: Trypanosoma cruzi

C/Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 09-Jul-2004

C/Accession: T30538

R/Bringing: F.; Vedrenne, C.; Cuvillier, A.; Parzy, D.; Baltz, D.; Tetaud, E.; Pays, E.,

Mol. Biochem. Parasitol. 94, 249-264, 1998

A/Title: Conserved organization of genes in trypanosomatids.

A/Reference number: Z16580; MUID:98418771; PMID:9747975

A/Accession: T30538

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-338 <BRI>

A/Cross-references: UNIPROT:O76230; EMBL:AF031927; NID:G3452217; PID:G3452219; PIDN:AC3

C/Genetics:

A/Gene: dnaJ

C/Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

F/4-70/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 90.5%; Score 38; DB 2; Length 338;

Best Local Similarity 75.0%; Pred. No. 9.3; Mismatches 0; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFER 8

DB 242 KHPHPR 249

RESULT 2

S71199

dnaJ protein homolog atj3 - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 28-Oct-1995 #sequence_revision 27-Feb-1997 #text_change 09-Jul-2004

C/Accession: S71199

R/Zhou, R.; Kroczyńska, B.; Miernik, J.A.

submitted to the EMBL Data Library, March 1995

A/Description: Atj3, an Arabidopsis thaliana homologue of the Escherichia coli DnaJ.

A/Reference number: S71199

A/Accession: S71199

A/Molecule type: mRNA

A/Residues: 1-420 <ZHO>

A/Cross-references: UNIPROT:Q42530; EMBL:U22340; NID:G1872162; PIDN:AAB49030.1; PID:G72727

C/Genetics:

A/Gene: atj3

C/Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

F/14-75/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 90.5%; Score 38; DB 2; Length 420;

Best Local Similarity 75.0%; Pred. No. 12; Mismatches 0; Indels 0; Gaps 0;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFER 8
 Db 260 KXHPQFER 267

RESULT 3

B82223
 hypothetical protein VC1253 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
 C/Species: Vibrio cholerae
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
 C/Accession: B82223
 R/Heldelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gilm, M.L.; Dodson, R.J.;
 Chaddson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.
 1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
 Nature 406, 477-483, 2000
 A/Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
 A/Reference number: A82035; MUID:20406833; PMID:10952301
 A/Accession: B82223
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-142 <HEI>
 A/Cross-references: UNIPROT:Q9XSK3; GB:AE004204; GB:AE003852; NID:g9655729; PIDN:AAF9441
 A/Experimental source: serogroup O1; strain N16961; biotype El Tor
 A/Genetic: VC1253
 A/Map position: 1.

Query Match
 Best Local Similarity 88.1%; Score 37; DB 2; Length 142;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KXHPQFER 8
 Db 86 KXHPQFER 93

RESULT 4

T36642
 hypothetical protein SCH35.17 - Streptomyces coelicolor
 C/Species: Streptomyces coelicolor
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C/Accession: T36642
 R/Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
 submitted to the EMBL Data Library, June 1999
 A/Reference number: Z21610
 A/Accession: T36642
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-189 <OUI>
 A/Cross-references: UNIPROT:Q9X8Y9; EMBL:AL078610; PIDN:GAB44410.1; GSPDB:GN00070; SCOEI
 A/Experimental source: strain A3(2)
 A/Genetic: SCOEI: SCH35.17

Query Match
 Best Local Similarity 88.1%; Score 37; DB 2; Length 189;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFER 7
 Db 105 KXHPQFER 111

RESULT 5

T00484
 hypothetical protein At2g35030 [imported] - Arabidopsis thaliana
 N/Alternate names: hypothetical protein F1913.26
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 09-Jul-2004
 C/Accession: T00484; G84763
 R/Rounsley, S.D.; Lin, X.; Keichum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
 submitted to the EMBL Data Library, April 1998
 A/Description: Arabidopsis thaliana chromosome II BAC F1913 genomic sequence.

A/Reference number: Z14160

A/Accession: T00484
 A/Status: translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-627 <ROU>

A/Cross-references: UNIPROT:Q64766; EMBL:AC004238; NID:g3033373; PID:g3033399
 A/Experimental source: cultivar Columbia
 R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shen, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;
 M.; Koop, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Unayam, L.; Tallon, L.
 Nature 402, 761-768, 1999
 A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A/Reference number: A84420; MUID:20083487; PMID:10617197
 A/Accession: G84763
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-627 <STO>
 A/Cross-references: GB:AB002093; NID:g3033399; PIDN:AA012843.1; GSPDB:GN00139
 A/Genetic: F1913.26; At2g35030
 A/Map position: 2

Query Match
 Best Local Similarity 88.1%; Score 37; DB 2; Length 627;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFER 7
 Db 590 KXHPQFER 596

RESULT 6

A53277
 hexokinase (EC 2.7.1.1) 1 - bovine
 C/Species: Bos primigenius taurus (cattle)
 C/Date: 10-Feb-1995 #sequence_revision 10-Feb-1995 #text_change 17-Mar-1999
 C/Accession: A53277; A3643
 R/Griffith, L.D.; Gelb, B.D.; Wheeler, D.A.; Davison, D.; Adams, V.; McCabe, E.R.
 Genomics 11, 1014-1024, 1991
 A/Title: Mammalian hexokinase 1: evolutionary conservation and structure to function an
 A/Reference number: A55277; MUID:92147096; PMID:1783373
 A/Accession: A55277
 A/Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tr
 A/Molecule type: mRNA
 A/Residues: 1-918 <GR>
 A/Experimental source: brain
 A/Note: sequence modified after extraction from NCBI backbone
 R/Griffith, L.D.; Macgregor, G.R.; Muzny, D.M.; Harter, J.; Cook, R.G.; McCabe, E.R.B.
 Biochem. Med. Metab. Biol. 41, 125-131, 1989
 A/Title: Synthesis and characterization of a bovine hexokinase 1 cDNA probe by mixed o11
 A/Accession: A43643; MUID:89247146; PMID:2719857
 A/Status: preliminary
 A/Molecule type: protein
 A/Residues: 11-51 <GR2>
 C/Superfamily: human hexokinase I; hexokinase homology
 C/Keywords: ATP; duplication; glycolysis; phosphotransferase
 F:479-907/Domain: hexokinase homology <HXK2>

Query Match
 Best Local Similarity 88.1%; Score 37; DB 2; Length 918;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KXHPQFER 8
 Db 866 KXHPQFER 873

RESULT 7

B71073
 probable malate dehydrogenase - Pyrococcus horikoshii

```

C:Species: Pyrococcus horikoshii
C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 12-Jul-2004
C:Accession: B71073
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohtoku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kuehida, N.; Oguchi
N. Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137; PMID:9679194
A:Accession: B71073
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-360 <KAW>
A:Cross-references: UNIPROT:O59028; GB:AP000005; NID:G3236132; PIDN:BA030380.1; PID:G325
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1277

Query Match      85.7% Score 36; DB 2; Length 360;
Best Local Similarity 62.5%; Pred. No. 24;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 KKHPOFER 8
       |||:|
Db      311 KKHPEFER 318

RESULT 8
AG3392
maleylpyruvate isomerase (EC 5.2.1.4) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AG3392
R:DeVecchio, V.D.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
. Mazur, M.; Goldstein, E.; Selkov, E.; Elzeir, P.H.; Hagius, S.; O'Callaghan, D.; Letese
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AJ3252; PMID:11756688
A:Accession: AG3392
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-215 <KUR>
A:Cross-references: UNIPROT:O8YGN3; GB:AE008917; PIDN:AAL52306.1; PID:G17983098; GSPDB:C
C:Genetics:
A:Gene: BME11125
A:Map position: 1
C:Keywords: cis-trans-isomerase

Query Match      83.3% Score 35; DB 2; Length 215;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 KKHPOFER 8
       |||:|
Db      191 KKHVFER 198

RESULT 9
S33312
dnaJ protease - leek (fragment)
C:Species: Allium porrum (leek)
C>Date: 06-Jan-1995 #sequence_revision 01-Sep-1995 #text_change 09-Jul-2004
C:Accession: S33312; #accession_change 09-Jul-2004
R:Beesoulle, J.J.
FEBS Lett. 323, 51-54, 1993
A:Title: Occurrence and sequence of a DnaJ protein in plant (Allium porrum) epidermal ce
A:Reference number: S33312; MUID:93265942; PMID:8495747
A:Accession: S33312
A:Molecule type: mRNA
A:Residues: 1-397 <BES>
A:Cross-references: UNIPROT:O03363; EMBL:X69436; NID:G16086; PIDN:CAA49211.1; PID:G16087
C:Genetics:

```

```

A:Gene: dnaJ
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
F.1-53/Domain: dnaJ amino-terminal homology (fragment) <DNJ>

Query Match      83.3% Score 35; DB 2; Length 397;
Best Local Similarity 62.5%; Pred. No. 42;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 KKHPOFER 8
       |||:|
Db      239 KKHPEFOR 246

RESULT 10
T40427
dnaJ related protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C:Accession: T40427
R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, March 1999
A:Reference number: Z21928
A:Accession: T40427
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-413 <SEB>
A:Cross-references: UNIPROT:O94657; EMBL:AL035655; PIDN:CAB38605.1; GSPDB:GN00067; SPDB:
A:Experimental source: strain 972h-; cosmid c405
C:Genetics:
A:Gene: SPDB:SPBC405.06
A:Map position: 2
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

Query Match      83.3% Score 35; DB 2; Length 413;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 KKHPOFER 8
       |||:|
Db      260 KKHVFER 267

RESULT 11
T23939
hypothetical protein R05H5.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T23939
R:McMurray, A.
submitted to the EMBL Data Library, March 1995
A:Reference number: Z19821
A:Accession: T23939
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-149 <MIT>
A:Cross-references: UNIPROT:O21763; EMBL:Z48795; PIDN:CAA8726.1; GSPDB:GN00020; CESP:RO
A:Experimental source: clone R05H5
C:Genetics:
A:Gene: CESP:R05H5.3
A:Map position: 2
A:Introns: 27/3; 98/2
C:Superfamily: Caenorhabditis elegans hypothetical protein C35H1.5

Query Match      81.0% Score 34; DB 2; Length 149;
Best Local Similarity 71.4%; Pred. No. 23;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 KKHPOFE 7
       |||:|
Db      57 KKHPEFE 63

RESULT 12

```

S74762
 hypothetical protein slr1619 - *Synechocystis* sp. (strain PCC 6803)
 C:Species: *Synechocystis* sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Jul-2004
 C:Accession: S74762
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
 O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*
 8.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S74762
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-249 <KAN>
 A:Cross-references: UNIPROT:P72897; EMBL:D90901; GB:AB001339; NID:g1651897
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match
 Best Local Similarity 81.0%; Score 34; DB 2; Length 249;
 Matches 5; Conservative 71.4%; Pred. No. 40; Indels 0; Gaps 0;

QY 1 KXHPQFE 7
 DB 220 KTHPEFE 226

RESULT 13
 E95998
 conserved hypothetical protein SMD21425 [imported] - *Sinorhizobium meliloti* (strain 1021)
 C:Species: *Sinorhizobium meliloti*
 C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 09-Jul-2004
 C:Accession: E95998
 R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan
 Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
 A:Title: The complete genome of the 1.683-kb pSymB megaplasmid from the N2-fixing endo
 A:Reference number: A95842; MUID:21396508; PMID:11481431
 A:Accession: E95998
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-258 <KUR>
 A:Cross-references: UNIPROT:Q92U84; GB:AL591985; PIDN:CAC49653.1; PID:g15141140; GSPDB:G
 A:Experimental source: strain 1021, megaplasmid pSymB
 R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abolaj, P.; Ampe, F.; Barloy-Hubler,
 J.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federici, N.A.; Fisher, R.F.;
 L.; Hyman, R.W.; Jones, T.
 Science 293, 668-672, 2001
 A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kies, E.; Komp, C.; Lelaure,
 hebaulic, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wellis, D.H.; Wong, K.; Yeh, K.
 A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
 A:Reference number: A96039; MUID:21368234; PMID:11474104
 A:Contents: annotation
 C:Genetics:
 A:Gene: SMD21425
 A:Genome: plasmid

Query Match
 Best Local Similarity 81.0%; Score 34; DB 2; Length 258;
 Matches 5; Conservative 71.4%; Pred. No. 41; Indels 0; Gaps 0;

QY 1 KXHPQFE 7
 DB 230 KSHPEFE 236

RESULT 14
 G75436
 conserved hypothetical protein - *Deinococcus radiodurans* (strain R1)
 C:Species: *Deinococcus radiodurans*
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C:Accession: G75436
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: G75436
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-300 <HT>
 A:Cross-references: UNIPROT:Q9RVC9; GB:AE001960; GB:AE000513; NID:g6458833; PIDN:AAF1067
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR1100
 A:Map position: 1
 C:Superfamily: Escherichia coli ycaA protein

Query Match
 Best Local Similarity 81.0%; Score 34; DB 2; Length 300;
 Matches 5; Conservative 71.4%; Pred. No. 48; Indels 0; Gaps 0;

QY 1 KXHPQFE 7
 DB 261 RAHPQFE 267

RESULT 15
 H69297
 conserved hypothetical protein AF0384 - *Archaeoglobus fulgidus*
 C:Species: *Archaeoglobus fulgidus*
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004
 C:Accession: H69297
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
 Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
 Nature 390, 364-370, 1997
 A:Authors: Overbeek, T.; Cotton, M.D.; Sprigge, T.; Artlich, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
 A:Reference number: A69250; MUID:98049343; PMID:9389475
 A:Accession: H69297
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-313 <KUR>
 A:Cross-references: UNIPROT:Q29863; GB:AE001078; GB:AE000782; NID:g2689401; PIDN:AA9085

Query Match
 Best Local Similarity 81.0%; Score 34; DB 2; Length 313;
 Matches 5; Conservative 71.4%; Pred. No. 51; Indels 0; Gaps 0;

QY 2 XHPQFE 8
 DB 267 YHPQFE 273

Search completed: March 2, 2005, 12:28:52
 Job time: 6.11382 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd

OM protein - protein search, using sw model

```
Run on:      March 2, 2005, 12:20:44 ; Search time 29.7236 Seconds
            (without alignments)
            ; 137.824 Million cell updates/sec
```

Title:	SEQ6
Perfect score:	42
Sequence:	1 kxhpgfer 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1612378 segs, 512079187 residues
Total number of hits satisfying chosen parameters: 1612378

```
Minimum DB seg length: 0
Maximum DB seg length: 2000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
1: uniprot_sprot:*
2: uniprot_trembl:
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	38	90.5	338	2	076230	076230 trypanosoma
2	38	90.5	338	2	09BIX8	09BIX8 trypanosoma
3	38	90.5	420	2	Q42530	Q42530 arabidopsis
4	37	88.1	142	2	Q9KSK3	Q9KSK3 vibrio chol
5	37	88.1	182	2	088XG7	088XG7 lactobacilli
6	37	88.1	189	2	Q9X8Y9	Q9X8Y9 streptomyces
7	37	88.1	335	2	Q7V101	Q7V101 prochloroc
8	37	88.1	366	2	Q9W083	Q9W083 drosophi
9	37	88.1	376	2	Q7X930	Q7X930 prunus avil
10	37	88.1	384	2	Q7VS20	Q7VS20 bordetella
11	37	88.1	384	2	Q7WEC2	Q7WEC2 bordetella
12	37	88.1	439	1	IDG1_DROYA	08mx-10 drosophila
13	37	88.1	557	2	08XRVO	08xrvo raletonia s
14	37	88.1	612	2	08TUD0	08tud0 methanosarc
15	37	88.1	627	2	064766	064766 arabidopsis
16	37	88.1	918	1	HXK1_BOVIN	P77555 bos taurus
17	37	88.1	5890	2	08IKR4	08ikr4 plasmodium
18	36	85.7	209	2	06LFX4	06lfx4 plasmodium
19	36	85.7	313	2	Q9P127	Q9p127 homo sapien
20	36	85.7	323	2	08LTT5	08lty5 vibriophag
21	36	85.7	360	1	MDH_PYRHO	059028 pyrococcus
22	36	85.7	551	2	Q98KJ2	Q98KJ2 rhizobium
23	35	83.3	215	2	08YGN3	08ygn3 bruceella m
24	35	83.3	215	2	08G181	08g181 bruceella s
25	35	83.3	223	2	08EG36	08eg36 shewanella
26	35	83.3	397	1	DNJ1_ALIPO	Q03363 allium por
27	35	83.3	410	2	Q73EB7	Q73EB7 bacillus ce
28	35	83.3	413	2	Q94657	Q94657 schizosacch
29	34	81.0	26	2	Q9ZG31	Q9ZG31 chlamydia t
30	34	81.0	48	2	Q48504	Q48504 lactococcus
31	34	81.0	123	2	Q9M674	Q9m674 abies alba

32	Q21763	Q21763	caenorhabditis
33	Q93N39	Q93N39	caenorhabditis
34	Q6R266	Q6R266	barcelona
35	P72897	P72897	synchocystis
36	Q64B27	Q64B27	uncultured
37	Q92U84	Q92U84	thizobium
38	YB00	YB00	deirna
39	Y384	Y384	arcpcu
40	Q8GCT37	Q8GCT37	solanum
41	Q95R57	Q95R57	drosophila
42	Q9VW98	Q9VW98	drosophila
43	Q20774	Q20774	caenorhabditis
44	O76224	O76224	tyrianosoma
45	DDB4	DDB4	human

ALIGNMENTS

ID	076230	PRELIMINARY;	PRT;	338 AA.
AC	076230.			
DT	01-NOV-1998	(TREMBLrel. 08, Created)		
DT	01-NOV-1998	(TREMBLrel. 08, Last sequence update)		
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)		
DE	Chaperone.			
GN	Name=DnaJ.			
OS	Trypanosoma cruzi.			
OC	Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.			
OX	NCBI_TaxID=5693;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CL.			
RA	MEDLINE=98418771; PubMed=9747975; DOI=10.1016/S0166-6851(98)00080-2;			
RA	Bringaud F., Vedemine C., Cuvellier A., Parzy D., Baltz D., Tetaud E.,			
RA	Pays E., Vengas J., Merlin G., Baltz T.;			
RT	"Conserved organization of genes in trypanosomatids.";			
RL	Mol. Biochem. Parasitol. 94:249-264(1998).			
DR	EMBL; AF031927; AAC32777.1; -.			
DR	PIR; T30538; T30538.			
DR	HSSP; P25685; 1HDJ.			
DR	GO; GO:0051082; P:unfolded protein binding; IEA.			
DR	GO; GO:0006457; P:protein folding; IEA.			
DR	InterPro; IPR002939; DnaJ_C.			
DR	InterPro; IPR001623; DnaJ_N.			
DR	InterPro; IPR008971; HSP40_DnaJ_pep.			
DR	InterPro; IPR003095; Hsp_DnaJ.			
DR	Pfam; PF00226; DnaJ_1.			
DR	Pfam; PF00566; DnaJ_C_1.			
DR	PRINTS; PR00625; DNaJPROTEIN.			
DR	SMART; SM00271; DnaJ_1.			
DR	PROSITE; PS00636; DNaJ_1; 1.			
DR	PROSITE; PS50076; DNaJ_2; 1.			
KM	Chaperone.			
SC	SEQUENCE 338 AA; 36535 MW; 95BA7EA791E2A19A CRC64;			
Query Match		90.5%;	Score 38;	DB 2; Length 338;
Best Local Similarity		75.0%;	Pred. No. 36;	
Matches	6; Conservative	2; Mismatches	0; Indels	0; Gaps
QY	1 KCHPOFER 8			
Db	242 KPHPFER 249			
RESULT 2				
Q9BIX8				
AC	Q9BIX8	PRELIMINARY;	PRT;	338 AA.
DT	01-JUN-2001	(TREMBLrel. 17, Created)		
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)		
DT	01-MAR-2004	(TREMBLrel. 26, Last annotation update)		

```

DE Co-chaperone protein.
GN Name=DnaJ-like;
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5693;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21570179; PubMed=11551903; DOI=10.1074/jbc.M102427200;
RA Salomon D., Montenegro-Lomeli M., Goldenberg S.;
RT "A DnaJ-like protein homologous to the yeast co-chaperone Sis1 (Ycj6p)
is involved in initiation of translation in Trypanosoma cruzi.";
RL J. Biol. Chem. 276:43970-43979(2001).
DR EMBL; AF345336; AAK19734.1; -.
DR HSSP; P25685; 1HDJ.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:006457; P:protein folding; IEA.
DR InterPro; IPR002938; DnaJ_C.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR008971; HSP40_DnaJ_pdp.
DR Pfam; PF00226; DnaJ; 1.
DR Pfam; PF01556; DnaJ_C; 1.
DR PRINTS; PR00625; DNJPROTEIN.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
KW Chaperone.
SQ SEQUENCE 338 AA; 36561 MW; 90A9E4F04508E55 CRC64;

Query Match
Best Local Similarity 90.5%; Score 38; DB 2; Length 338;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFER 8
Db 242 KXHPQFER 249
|:|:|:|:|
|:|:|:|:|

RESULT 3
ID Q42530 PRELIMINARY; PRT; 420 AA.
AC Q42530;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE DnaJ homolog;
GN Name=atj;
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20027448; PubMed=10557255;
RA Zhou R., Koczyska B., Miernyk J.A.;
RT "Atj3 (Accession No. U22340), an Arabidopsis thaliana J-Protein
homologous to Saccharomyces cerevisiae Ydj1p. (PGR99-162).";
RL EMBL; U22340; AAB49030.1; -.
DR EMBL; U22340; AAB49030.1; -.
DR PIR; S71199; S71199.
DR HSSP; P25685; 1HDJ.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:006457; P:protein folding; IEA.
DR InterPro; IPR002939; DnaJ_C.
DR InterPro; IPR001305; DnaJ_CXKXGKG.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR008971; HSP40_DnaJ_pdp.
DR InterPro; IPR003095; Hsp_DnaJ.
DR Pfam; PF00226; DnaJ; 1.
DR Pfam; PF01556; DnaJ_C; 1.
DR Pfam; PF00684; DnaJ_CXKXGKG; 1.
DR PRINTS; PR00625; DNJPROTEIN.

```

```

DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
DR PROSITE; PS00637; DnaJ_CXKXGKG; 1.
KW Chaperone; Repeat.
SQ SEQUENCE 420 AA; 46444 MW; C4C12848F61AD445 CRC64;

Query Match
Best Local Similarity 90.5%; Score 38; DB 2; Length 420;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFER 8
Db 260 KXHPQFER 267
|:|:|:|:|
|:|:|:|:|

RESULT 4
ID Q9KSK3 PRELIMINARY; PRT; 142 AA.
AC Q9KSK3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Hypothetical protein VC1253.
GN OrderedLocustNames=VC1253;
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301; DOI=10.1038/35020000;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tetelin H., Richardson D.L.,
RA Ermolaeva M.D., Vamathevan J.J., Bass S., Qin H., Dragol I.,
RA Sellers P., McDonald L.A., Utterback T.R., Fleischmann R.D.,
RA Niernan W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R.,
RA Mekalanos J.J., Venter J.C., Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004204; AAF94412.1; -.
DR PIR; B82223; B82223.
DR TIGR; VC1253; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 142 AA; 16872 MW; AEE0459706E2F1CA CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 2; Length 142;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KXHPQFER 8
Db 86 KXHPQFER 93
|:|:|:|:|
|:|:|:|:|

RESULT 5
ID Q88XG7 PRELIMINARY; PRT; 182 AA.
AC Q88XG7;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Hypothetical protein lp_1238.
GN OrderedLocustNames=lp_1238;
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RC STRAIN=NCIMB 8826 / WCFS1;
RX MEDLINE=22480296; PubMed=1256566; DOI=10.1073/pnas.0337704100;
RA Kleebebezen M., Beekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Tarchini R., Peters S.A., Sandring H.M.,
RA Fliers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hoffer S.M., Nierop Groot M.N., Kerkhoven R., De Vries M., Ureling B.,
RA De Vos W.M., Stezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCFS1."
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995(2003).
RW EMBL; AL935235; CADD3746.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 182 AA; 21722 MW; 0C124611A43D8CF CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 2; Length 182;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFE 7
DB 35 KXHPQFE 41

RESULT 6
QX8Y9 PRELIMINARY; PRT; 189 AA.
AC QX8Y9;
DT 01-NOV-1999 (TRMBLrel. 12, Created)
DT 01-NOV-1999 (TRMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TRMBLrel. 26, Last annotation update)
DE Putative lipoprotein.
GN ORFNames=SC35.17;
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycetaceae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxId=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=2196410; PubMed=12000953; DOI=10.1038/41741a;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleeer H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Kleeer T., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleeer T., Larke L., Murphy L.D., Oliver K., O'Neill S.,
RA Radniewitch E., Rajandream M.A., Rutherford K.M., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J.R., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147(2002).
DR EMBL; AL939117; CAB44410.1; -.
DR PIR; T36642.
KW Complete proteome; Lipoprotein.
SQ SEQUENCE 189 AA; 19772 MW; 82D231E01CAC3B57 CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 2; Length 189;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFE 7
DB 105 KXHPQFE 111

RESULT 7
Q7V101 PRELIMINARY; PRT; 335 AA.
AC Q7V101;
DT 01-OCT-2003 (TRMBLrel. 25, Created)
DT 01-OCT-2003 (TRMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TRMBLrel. 26, Last annotation update)
DE Biotin synthase (EC 2.8.1.6).

```

```

GN Name=BIOB; OrderedLocustNames=PMW1093;
OS Prochlorococcus marinus subsp. pastoris (strain CCMP 1378 / MED4).
OC Bacteria; Cyanobacteria; Prochlorales; Prochlorococaceae;
OC Prochlorococcus.
OX NCBI_TaxId=59919;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22825698; PubMed=12917642; DOI=10.1038/nature01947;
RA Rocap G., Larimer F.W., Lamerdin J.E., Malfatti S., Chain P.,
RA Ahlgren N.A., Arriaga A., Coleman M., Hauser L., Hess W.R.,
RA Johnson Z.I., Land M.L., Lindell D., Post A.F., Regala W., Shah M.,
RA Shaw S.L., Steglich C., Sullivan M.B., Ting C.S., Tolonen A.,
RA Webb B.A., Zinser E.R., Chisholm S.W.;
RT "Genome divergence in two Prochlorococcus ecotypes reflects oceanic
RT niche differentiation."
RL Nature 424:1042-1047(2003).
RW EMBL; BX572092; CAE19552.1; -.
DR HSSP; P12996; 1R30.
DR GO; GO:0004076; P:biotin synthase activity; IEA.
DR GO; GO:0005506; P:iron ion binding; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0009102; P:biotin biosynthesis; IEA.
DR InterPro; IPR010722; BATS.
DR InterPro; IPR002684; Biotin synth.
DR InterPro; IPR006638; Elp3/Mi18/Ni18.
DR InterPro; IPR007197; Radical_SAM.
DR Pfam; PF06968; BATS; 1.
DR Pfam; PF04055; Radical_SAM; 1.
DR SMART; SM00729; Elp3; 1.
DR TIGRFAMs; TIGR00433; biOB; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 335 AA; 37719 MW; EB93FB56044E04D CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 2; Length 335;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFE 7
DB 79 KXHPQFE 85

RESULT 8
Q9W083 PRELIMINARY; PRT; 366 AA.
AC Q9W083;
DT 01-MAY-2000 (TRMBLrel. 13, Created)
DT 01-MAY-2000 (TRMBLrel. 13, Last sequence update)
DT 01-MAR-2004 (TRMBLrel. 26, Last annotation update)
DE CG12020-PA.
GN ORFNames=CG12020;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OX NCBI_TaxId=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gockyne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazek R.G., Chape M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Neilson C.R., Gaber G.L.,
RA Abell J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Balow R.W., Basu P.V., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck U., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davaport L.B., Davies P.,
RA de Pablo B., Deitcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

```

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.U., Mei M.H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kechum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Miletina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.Y., Waasman D.A., Weinstein G.M., Weisenbach J.,
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RA Science 287:2185-2195(2000).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=22426065; PubMed=12537568;
 RX Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
 RA George R.A., Hoskins R.A., Laverly T., Munz D.M., Nelson C.R.,
 RA Paclet J.M., Park S., Pfeiffer B.D., Richards S., Sodegren E.J.,
 RA Svitek R., Tabor P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
 RA "Finishing a whole-genome shotgun: Release 3 of the *Drosophila*
 RA *melanogaster* euchromatic genome sequence.";
 RA Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
 RL [3]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=22426070; PubMed=12537573;
 RX Celniker J.S., Bergman C.M., Kronmiller B., Carlson J., Svitek R.,
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
 RA Ashburner M., Celniker S.E.;
 RA "The transposable elements of the *Drosophila melanogaster* euchromatin:
 RA a genomic perspective.";
 RA Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
 RL [4]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=22426069; PubMed=12537572;
 RX Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.B.,
 RA Smith C.D., Tupy J.L., Whitfield E.U., Bayraktaroglu L., Bertman B.P.,
 RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
 RA Harris N.L., Richter J., Russo S., Schroeder A.U., Shu S.Q.,
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 RA Lewis S.E.;
 RA "Annotation of the *Drosophila melanogaster* euchromatic genome: a
 RA systematic review.";
 RA Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
 RL [5]
 RN SEQUENCE FROM N.A.
 RP FlyBase;
 RG Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RL [6]
 RN SEQUENCE FROM N.A.
 RP FlyBase;
 RG Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AE003472; AAF47572.1; -
 DR HSSP; P25685; 1HDJ.
 DR InRAct; OSW083; -
 DR FlyBase; FBgn0035273; CG12020.
 DR GO; GO:0051082; F:unfolded protein binding; IEA.
 DR GO; GO:0006457; P:protein folding; IEA.
 DR InterPro; IPR002939; DnaJ_C.

DR InterPro; IPR001623; DnaJ_N.
 DR InterPro; IPR008971; HSP40_DnaJ_pap.
 DR Pfam; PF00226; DnaJ_1.
 DR Pfam; PF01556; DnaJ_C1.
 DR SMART; SM00271; DnaJ; 1.
 DR PROSITE; PS50076; DnaJ_2; 1.
 SQ SEQUENCE 366 AA; 42459 MW; 4D9A86AFC446D62 CRC64;
 QY Query Match 88.1%; Score 37; DB 2; Length 366;
 Db Best Local Similarity 75.0%; Pred. No. 62;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KXHPFER 8
 Db 237 KHPDFER 244
 ID Q7X930 PRELIMINARY; PRT; 376 AA.
 AC Q7X930;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
 DE S-locus F-box protein 3.
 DE Name=SF83;
 GN Prunus avium (Cherry).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eustosids I; Rosales; Rosaceae; Amygaloideae; Prunus.
 OX NCBI_TaxID=42229;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22763733; PubMed=12881505;
 RA Yanane H., Ikeda K., Ushijima K., Sassa H., Tao R.;
 RA "A pollen-expressed gene for a novel protein with an F-box motif that
 RA is very tightly linked to a gene for S-RNase in two species of cherry,
 RA *Prunus cerasus* and *P. avium*.";
 RA Plant Cell Physiol. 44:764-769(2003).
 RL EMBL; AB096857; BAC0148.1; -
 DR InterPro; IPR001810; F-box.
 DR InterPro; IPR006527; F-box_assoc.1.
 DR InterPro; IPR011043; Gal_oxid_central.
 DR InterPro; IPR000504; RNA_rec_mot.
 DR Pfam; PF00646; F-box; 1.
 DR SMART; SM00256; FBOX; 1.
 DR TIGRFAMs; TIGR01640; F_box_assoc.1; 1.
 DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN.1.
 SQ SEQUENCE 376 AA; 43962 MW; 607633857936587 CRC64;
 QY Query Match 88.1%; Score 37; DB 2; Length 376;
 Db Best Local Similarity 85.7%; Pred. No. 63;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 XHPFER 8
 Db 63 HHPFER 69
 ID Q7VS20 PRELIMINARY; PRT; 384 AA.
 AC Q7VS20;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN OrderedCusNames=BP0636;
 OS Bordetella pertussis.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Alcaligenaceae; Bordetella.
 OX NCBI_TaxID=520;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=phoma I / ATCC BAA-589 / NCTC 13251;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdeno-Tarrega A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Felwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
 RA Lealher S., Moule S., Norbertczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrall B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of *Bordetella pertussis*,
 RT *Bordetella parapertussis* and *Bordetella bronchiseptica*.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640412; CAE44962.1; -;
 KW Complete proteome; Hypochemical protein.
 SQ SEQUENCE 384 AA; 39424 MW; 4A23968B197193F2 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 384;
 Best Local Similarity 85.7%; Pred. No. 65;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPOFER 8
 Db 130 GHPOFER 136

RESULT 11

07WEC2 PRELIMINARY; PRT; 384 AA.

AC 07WEC2;
 DT 01-OCT-2003 (TRMBLrel. 25, Created)
 DT 01-OCT-2003 (TRMBLrel. 25, Last sequence update)
 DE Hypochemical protein.
 GN OrderedCusNames=BA4712;
 OS *Bordetella bronchiseptica* (Alcaligenes bronchiseptica).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC *Alcaligenaceae*; *Bordetella*.
 OC NCBI_TaxID=518;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RSO / ATCC BAA-589;
 RX MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
 RA Parkhill J., Sebahia M., Preston A., Murphy L.D., Thomson N.R.,
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
 RA Cerdeno-Tarrega A.-M., Temple L., James K.D., Harris B., Quail M.A.,
 RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
 RA Felwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
 RA Lealher S., Moule S., Norbertczak H., O'Neill S., Ormond D., Price C.,
 RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
 RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
 RA Unwin L., Whitehead S., Barrall B.G., Maskell D.J.;
 RT "Comparative analysis of the genome sequences of *Bordetella pertussis*,
 RT *Bordetella parapertussis* and *Bordetella bronchiseptica*.";
 RL Nat. Genet. 35:32-40(2003).
 DR EMBL; BX640451; CAE35075.1; -;
 KW Complete proteome; Hypochemical protein.
 SQ SEQUENCE 384 AA; 39467 MW; 779AEC7C42AB9038 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 384;
 Best Local Similarity 85.7%; Pred. No. 65;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPOFER 8
 Db 130 GHPOFER 136

RESULT 12

Q8XRVO PRELIMINARY; PRT; 557 AA.

IDG1 DROYA
 ID IDG1 DROYA STANDARD; PRT; 439 AA.
 AC O8XR40;
 DT 29-MAR-2004 (Rel. 43, Created)
 DT 29-MAR-2004 (Rel. 43, Last sequence update)
 DT 25-OCT-2004 (Rel. 45, Last annotation update)
 DE Chitinase-like protein Idgf1 precursor (imaginal disk growth factor
 DE protein 1).
 GN Name=Idgf1.
 OS *Drosophila yakuba* (Fruit fly).
 OS *Eukaryota*; *Metazoa*; *Arthropoda*; *Insecta*; *Pterygota*;
 OC *Neoptera*; *Endopterygota*; *Diptera*; *Brachycera*; *Muscomorpha*;
 OC *Ephydroidea*; *Drosophilidae*; *Drosophila*.
 OC NCBI_TaxID=7245;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H14021-0261.0;
 RX MEDLINE=22226651; PubMed=12242232;
 RA Zurovcova M., Ayala F.J.;
 RT "Polymorphism patterns in two tightly linked developmental genes,
 RT *Idgf1* and *Idgf3*, of *Drosophila melanogaster*.";
 RL Genetics 162:177-188(2002).
 CC -1- FUNCTION: Cooperates with insulin-like peptides to stimulate the
 CC proliferation, polarization and motility of imaginal disk cells.
 CC May act by stabilizing the binding of insulin-like peptides to its
 CC receptor through a simultaneous interaction with both molecules to
 CC form a multiprotein signaling complex (By similarity).
 CC -1- SUBCELLULAR LOCATION: Secreted. It is probably transported to
 CC target tissues via hemolymph (By similarity).
 CC -1- PTM: Glycosylated (By similarity).
 CC -1- MISCELLANEOUS: Lacks the typical Glu active site in position 150
 CC that is replaced by a Gln residue, preventing the hydrolase
 CC activity. Its precise function remains unclear.
 CC -1- SIMILARITY: Belongs to the glycosyl hydrolase 18 family. IDGP
 CC subfamily.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AF394712; AAM69644.1; -;
 DR HSSP; O96665; JUND.
 DR Flybase; FBgn0044136; Dyak\ldgf1.
 DR InterPro; IPR001223; Glyco_hydro_18.
 DR Pfam; PF00704; Glyco_hydro_18; 1.
 DR ProDom; PD000471; Chitinase_II; 1.
 DR SMART; SM00636; Glyco_18; 1.
 KW Developmental protein; Glycoprotein; Signal.
 FT SIGNAL 1 20
 FT CHAIN 1 439 By similarity.
 FT DISULFID 26 53 Chitinase-like protein Idgf1.
 FT DISULFID 340 423 By similarity.
 FT CARBOHYD 122 122 N-linked (GlcNAc...) (Potential).
 FT CARBOHYD 218 218 N-linked (GlcNAc...) (By similarity).
 FT CARBOHYD 346 346 N-linked (GlcNAc...) (Potential).
 SQ SEQUENCE 439 AA; 49244 MW; EDE16BD82A1B9E CRC64;

Query Match 88.1%; Score 37; DB 1; Length 439;
 Best Local Similarity 85.7%; Pred. No. 75;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 XHPOFER 7
 Db 227 KXHPFER 233

RESULT 13

Q8XRVO PRELIMINARY; PRT; 557 AA.

AC O8XRV0;
 DT 01-MAR-2002 (Tremblrel. 20, Created)
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 26, Last annotation update)
 DE PROBABLE TREHALOSE-6-PHOSPHATE SYNTHASE (ALPHA, ALPHA-TREHALOSE-
 PHOSPHATE SYNTHASE UDP-FORMING) PROTEIN (EC 2.4.1.15).
 GN Name=RS01697; OrderedLocustNames=RS0731.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OG Plasmid megaplasmid.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Burkholderiaceae; Ralstonia.
 NC NCB1_TaxID=305;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM1100;
 RX MEDLINE=2161879; PubMed=11823852; DOI=10.1038/415497a;
 RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
 RA Chancelier M., Choigne N., Claudel-Renard C., Cunac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Siguler P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weissbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
 RL Nature 415:497-502(2002).
 DR EMBL; AL646080; CAD17882.1; -.
 DR HSSP; P31677; 1G25.
 DR GO; GO:0003825; F:alpha,alpha-trehalose-phosphate synthase (U. . .; IEA.
 DR GO; GO:0005992; P:trehalose biosynthesis; IEA.
 DR PFam; PF00982; Glyco_transf_20; 1.
 DR PROSITE; PS00276; CHANNEL_COLICIN; UNKNOWN_1.
 KW Complete proteome.
 SQ SEQUENCE 557 AA; 61659 MW; 4631D135F80B62E7 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 557;
 Best Local Similarity 85.7%; Pred. No. 98;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOPER 8
 DB 40 RHPOPER 46

RESULT 14
 Q8TUH0 PRELIMINARY; PRT; 612 AA.
 AC Q8TUH0;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE GLYCYL-CRNA synthetase.
 GN Name=glys; OrderedLocustNames=MA0097;
 OS Methanosarcina acetivorans.
 OC Archaea; Euryarchaeota; Methanomicetobria; Methanosarcinales;
 OC Methanosarcinaceae; Methanosarcina.
 NC NCB1_TaxID=2214;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C2A / ATCC 35395 / DSM 2834;
 RX MEDLINE=21929760; PubMed=11932238; DOI=10.1101/gr.223902;
 RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., Macdonald P., Brown A.,
 RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Altner D., Brown R.,
 RA Allen N., Naylor J., Stange-Thomann N., Dearellano K., Johnson R.,
 RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
 RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
 RA Hedderich R., Ingram-Smith C., Kueltner H.C., Krzycki J.A.,
 RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
 RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
 RA Ferry J.G., Jarell K.F., Jing H., Macario A.J.U., Paulsen I.T.,
 RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
 RA Metcalf W.W., Birren B.;
 RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
 and physiological diversity."
 RL Genome Res. 12:532-542(2002).

CC -I- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
 DR EMBL; AE010667; AA003551.1; -.
 DR HSSP; P56206; 1AT1.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004820; F:glycine-cRNA ligase activity; IEA.
 DR GO; GO:0006425; P:glycyl-cRNA aminoacylation; IEA.
 DR GO; GO:0006412; P:protein biosynthesis; IEA.
 DR PFam; PF03129; tRNA-synt_2b; 1.
 DR PFam; PF00587; tRNA-synt_2b; 1.
 DR PRINTS; PR01043; TRNASYNTHLY.
 DR TIGRPFAM; TIGR00389; glys_dimeric; 1.
 DR PROSITE; PS00862; AA_TRNA_LIGASE_II; 1.
 KW Complete proteome.
 SQ SEQUENCE 612 AA; 69165 MW; A12AF972927C8B55 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 612;
 Best Local Similarity 75.0%; Pred. No. 11e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KXHPOPER 8
 DB 222 KXHPOPER 229

RESULT 15
 ID 064766 PRELIMINARY; PRT; 627 AA.
 AC 064766;
 DT 01-AUG-1998 (Tremblrel. 07, Created)
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
 DE Hypothetical protein At2g35030.
 GN Name=At2g35030;
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophytes; Magnoliophyta; eudicotyledons; core eudicotyledons;
 OC Eusteroideae II; Brassicales; Brassicaceae; Arabidopsie.
 NC NCB1_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
 RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
 RA Somerville C.R., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Town C.D., Kaul S.;
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC004238; AAC12843.1; -.
 DR PIR; T00484; T00484.
 DR InterPro; IPR002885; PPR.
 DR InterPro; IPR008940; Prenyl_trans.
 DR InterPro; IPR008941; TPR-1like.
 DR PFam; PF01535; PPR; 12.
 DR TIGRPFAM; TIGR00756; PPR; 10.
 KW Hypothetical protein.
 SQ SEQUENCE 627 AA; 71409 MW; 27BDA2EC168AF949 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 627;
 Best Local Similarity 85.7%; Pred. No. 1.1e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPOPER 7
 DB 590 KXHPOPER 596

Search completed: March 2, 2005, 12:44:17
 Job time: 30.7236 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 36.0325 Seconds
(without alignments) 85.869 Million cell updates/sec

Title: SEQ6
Perfect score: 42
Sequence: 1 kxhpqfer 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq.16Dec04:*
1: geneseqp1808:*
2: geneseqp1908:*
3: geneseqp2008:*
4: geneseqp20018:*
5: geneseqp20028:*
6: geneseqp20038:*
7: geneseqp20038:*
8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	482	7	ADC95241
2	37	88.1	11	2	AAR52694
3	37	88.1	11	2	AAR52693
4	37	88.1	12	2	AAR52697
5	37	88.1	366	4	ABBS7854
6	36	85.7	231	7	ADBS5422
7	36	85.7	313	4	AAB75608
8	36	85.7	359	8	ADBS1924
9	36	85.7	360	8	ADN18635
10	36	85.7	598	8	ABM83512
11	35	83.3	378	8	ADN25797
12	35	83.3	455	8	ADBS23840
13	34	81.0	8	2	AAW59212
14	34	81.0	8	4	AAH35433
15	34	81.0	8	4	AAH35442
16	34	81.0	8	4	AAH35442
17	34	81.0	8	5	AAO19065
18	34	81.0	8	5	ABBS7464
19	34	81.0	8	5	ABBS7486
20	34	81.0	8	5	ABG73584
21	34	81.0	8	6	AAE37230
22	34	81.0	8	6	ABP60361
23	34	81.0	8	6	ABP60368
24	34	81.0	8	7	ADA09808
25	34	81.0	8	7	ADB84588

ALIGNMENTS

26	34	81.0	8	7	ADB85500	ADB85500 Streptococcus
27	34	81.0	8	7	ADD29930	ADD29930 Anticodon
28	34	81.0	8	8	ADP06951	ADP06951 Streptococcus
29	34	81.0	8	8	ADP06952	ADP06952 Protein/p
30	34	81.0	8	8	ADBS20244	ADBS20244 Streptococcus
31	34	81.0	8	8	ADBS20244	ADBS20244 Streptococcus
32	34	81.0	9	5	ABG31054	ABG31054 Peptide p
33	34	81.0	10	4	AA976652	AA976652 Influenza
34	34	81.0	10	4	AA976652	AA976652 Streptococcus
35	34	81.0	10	5	AAU80475	AAU80475 Peptide S
36	34	81.0	10	5	ABP56623	ABP56623 C-termina
37	34	81.0	10	8	ADP11064	ADP11064 Streptococcus
38	34	81.0	10	8	ADN16967	ADN16967 Human res
39	34	81.0	10	8	ADBS26489	ADBS26489 Streptococcus
40	34	81.0	11	6	AAE38373	AAE38373 Epitope t
41	34	81.0	11	6	ADBS20228	ADBS20228 Streptococcus
42	34	81.0	13	2	AAE78379	AAE78379 Synthetic
43	34	81.0	19	6	ABG74881	ABG74881 Bacteriophage
44	34	81.0	19	6	ABG74882	ABG74882 Bacteriophage
45	34	81.0	19	6	ABG74880	ABG74880 Bacteriophage

RESULT 1
ADC95241
ID ADC95241 standard; protein; 482 AA.
XX
AC ADC95241;
XX
DT 01-JUN-2004 (first entry)
XX
DE E. faecium protein sequence SEQ ID 4868.
XX
KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
XX abdominal-pelvic infection.
XX
OS Enterococcus faecium.
XX
PN US683275-B1.
XX
PD 24-JUN-2003.
XX
PF 30-JUN-1998; 98US-00107532.
XX
PR 02-JUL-1997; 97US-0051571P.
XX
PR 14-MAY-1998; 98US-0085598P.
XX
(GENO-) GENOME THERAPEUTICS CORP.
XX
PI Doucette-Stamm LA, Bush D;
XX
PI WPI; 2003-799836/75.
XX
DR N-PSDB; ADC91587.
XX
PT New isolated nucleic acid derived from Enterococcus faecium encoding an
XX Enterococcus faecium polypeptide useful for detection, prevention and
XX treatment of a pathological condition resulting from a bacterial
XX infection.
XX
PS Example 1; SEQ ID NO 4868; 243pp; English.
XX
CC The invention relates to an isolated nucleic acid derived from
XX Enterococcus faecium encoding an Enterococcus faecium polypeptide having
XX one of 10 fully defined sequences given in the (or comprising 40
XX sequential nucleotides chosen from any of the nucleic acids, its
XX complement or sequences hybridising to it). Also included are a
XX recombinant vector comprising the nucleic acid operably linked to
XX transcription regulatory element, a cell comprising the vector and a
XX single-stranded probe comprising the nucleic acid. The nucleic acids are
XX chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
XX The nucleic acids is useful for diagnosing pathological conditions

CC resulting from E. faecium bacterial infection (e.g. urinary tract
 CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
 CC infection), and for screening drugs such as agonists and antagonists. The
 CC nucleic acid is useful for recombinant production of *Candida albicans* -
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions
 CC and vaccines containing the nucleic acid are useful for preventing or
 CC treating *Enterococcus faecium* infections. The present sequence represents
 CC one if the disclosed E. faecium proteins.
 XX
 SQ Sequence 482 AA;

Query Match 92.9%; Score 39; DB 7; Length 482;
 Best Local Similarity 75.0%; Pred. No. 78;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKHQPER 8
 DB 409 KKHQPER 416

RESULT 2
 AAR52694
 ID AAR52694 standard; protein; 11 AA.

AC AAR52694;
 XX
 DT 10-JAN-1995 (first entry)
 XX
 DE pASK46-p14XH encoded C-terminal streptavidin-binding sequence.

KM Streptavidin binding peptide; fusion protein; immunoglobulin; Ig;
 KW heavy chain variable region; affinity chromatography; purification;
 XX peptide tag.

OS Synthetic.

FT Key Location/Qualifiers
 FT Peptide 4..11
 FT /note="streptavidin-binding octapeptide fused to C-terminus of VH chain"

GB2272698-A.

25-MAY-1994.

PF 01-NOV-1993; 93GB-00022501.

PR 03-NOV-1992; 92DE-04237113.

PA (BIOA-) INST BIOANALYTIK GEMEINNUTZIGE GMBH.

PI Skerra A, Schmidt T;

DR WPI; 1994-153484/19.

DR N-PSDB; AAQ62669.

PT New fusion peptide(s) - have easily controlled binding properties and are
 PT capable of binding to streptavidin.
 XX
 PS Disclosure; Page 11; 53pp; English.

XX
 CC Six derivatives of pASK46 (a plasmid for expression of the D1.3FV
 CC fragment in E.coli) were produced which encode 4 different peptides at
 CC the C-terminus of VH and 2 peptides at the C-terminus of VL of the D1.3FV
 CC fragment. The peptides fused to the C-terminus are all examples of
 CC streptavidin-binding peptides corresponding to a generic formula (see
 CC AAR52698). The peptides do not interfere with the protein function but
 CC facilitate purification by conferring streptavidin-binding properties on
 CC the fusion protein

SQ Sequence 11 AA;

Query Match 88.1%; Score 37; DB 2; Length 11;

Best Local Similarity 85.7%; Pred. No. 3;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHHQPER 8
 DB 5 XHHQPER 11

RESULT 3
 AAR52693

ID AAR52693 standard; protein; 11 AA.

AC AAR52693;

DT 10-JAN-1995 (first entry)

DE pASK46-p14XH encoded C-terminal streptavidin-binding sequence.

KM Streptavidin binding peptide; fusion protein; immunoglobulin; Ig;
 KW heavy chain variable region; affinity chromatography; purification;
 XX peptide tag.

OS Synthetic.

FT Key Location/Qualifiers
 FT Peptide 4..11
 FT /note="streptavidin-binding octapeptide fused to C-terminus of VH chain"

GB2272698-A.

25-MAY-1994.

PF 01-NOV-1993; 93GB-00022501.

PR 03-NOV-1992; 92DE-04237113.

PA (BIOA-) INST BIOANALYTIK GEMEINNUTZIGE GMBH.

PI Skerra A, Schmidt T;

DR WPI; 1994-153484/19.

DR N-PSDB; AAQ62668.

PT New fusion peptide(s) - have easily controlled binding properties and are
 PT capable of binding to streptavidin.
 XX
 PS Disclosure; Page 11; 53pp; English.

XX
 CC Six derivatives of pASK46 (a plasmid for expression of the D1.3FV
 CC fragment in E.coli) were produced which encode 4 different peptides at
 CC the C-terminus of VH and 2 peptides at the C-terminus of VL of the D1.3FV
 CC fragment. The peptides fused to the C-terminus are all examples of
 CC streptavidin-binding peptides corresponding to a generic formula (see
 CC AAR52698). The peptides do not interfere with the protein function but
 CC facilitate purification by conferring streptavidin-binding properties on
 CC the fusion protein

SQ Sequence 11 AA;

Query Match 88.1%; Score 37; DB 2; Length 11;
 Best Local Similarity 85.7%; Pred. No. 3;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHHQPER 8
 DB 5 XHHQPER 11

RESULT 4
 AAR52697

ID AAR52697 standard; protein; 12 AA.

```

AC  AAR52697;
XX
XX  10-JAN-1995 (first entry)
XX
XX  pASK46-gliXL encoded C-terminal streptavidin-binding sequence.
DE
XX  pASK46-gliXL encoded C-terminal streptavidin-binding sequence.
XX
XX  Streptavidin binding peptide; fusion protein; immunoglobulin; Ig;
XX  light chain variable region; affinity chromatography; purification;
XX  peptide tag.
XX
XX  Synthetic.
OS
XX
XX  Key
XX  Peptide
XX
XX  Location/Qualifiers
XX  5..12
XX  /note="streptavidin-binding octapeptide fused to C-
XX  terminus of VL chain"
XX
XX  GB2272698-A.
XX
XX  25-MAY-1994.
XX
XX  01-NOV-1993; 93GB-00022501.
XX
XX  03-NOV-1992; 92DB-04237113.
XX
XX  (BIOA-) INST BIOANALYTIK GEMEINNUTZIGE GMBH.
XX
XX  Skerra A, Schmidt T;
XX
XX  WPI; 1994-153484/19.
XX
XX  N-PSDB; AAQ62672.
XX
XX  New fusion peptide(s) - have easily controlled binding properties and are
XX  capable of binding to streptavidin.
XX
XX  Disclosure; Page 11; 53pp; English.
XX
XX  Six derivatives of pASK46 (a plasmid for expression of the D1.3Fv
XX  fragment in E.coli) were produced which encode 4 different peptides at
XX  the C-terminus of VH and 2 peptides at the C-terminus of VL of the D1.3Fv
XX  fragment. The peptides fused to the C-terminus are all examples of
XX  streptavidin-binding peptides corresponding to a generic formula (see
XX  CC AAR52698). The peptides do not interfere with the protein function but
XX  facilitate purification by conferring streptavidin-binding properties on
XX  the fusion protein
XX
XX  Sequence 12 AA;
SQ
XX
XX  Query Match
XX  Best Local Similarity 88.1%; Score 37; DB 2; Length 12;
XX  Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX  2 XHPOPER 8
XX  :|||||
XX  6 RHPQPER 12
XX
XX  RESULT 5
XX  ID ABB57854
XX  ABB57854 standard; protein; 366 AA.
XX
XX  ABB57854;
XX
XX  26-MAR-2002 (first entry)
XX
XX  Drosophila melanogaster polypeptide SEQ ID NO 354.
XX
XX  Drosophila; developmental biology; cell signaling; insecticide;
XX  pharmaceutical.
XX
XX  Drosophila melanogaster.
XX
XX  WO200171042-A2.
XX

```

```

XX
XX  27-SEP-2001.
XX
XX  23-MAR-2001; 2001WO-US009231.
XX
XX  23-MAR-2000; 2000US-0191637P.
XX
XX  11-JUL-2000; 2000US-00614150.
XX
XX  (PEKE ) PE CORP NY.
XX
XX  Venter JC, Adams M, Li PWD, Myers EW;
XX
XX  WPI; 2001-656860/75.
XX
XX  N-PSDB; ABL01957.
XX
XX  New isolated nucleic acid detection reagent for detecting 1000 or more
XX  genes from Drosophila and for elucidating cell signaling and cell-cell
XX  interactions.
XX
XX  Disclosure; SEQ ID NO 354; 21pp + Sequence Listing; English.
XX
XX  The invention relates to an isolated nucleic acid detection reagent
XX  capable of detecting 1000 or more genes from Drosophila. The invention is
XX  useful in developmental biology and in elucidating cell signaling and
XX  cell-cell interactions in higher eukaryotes for the development of
XX  insecticides, therapeutics and pharmaceutical drugs. The invention
XX  discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX  sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
XX  ABB72072). The sequence data for this patent did not form part of the
XX  printed specification, but was obtained in electronic format directly
XX  from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX  Sequence 366 AA;
SQ
XX
XX  Query Match
XX  Best Local Similarity 88.1%; Score 37; DB 4; Length 366;
XX  Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX  1 XHPOPER 8
XX  :|||||
XX  237 KHPQPER 244
XX
XX  RESULT 6
XX  ID ADB65422
XX  ADB65422 standard; protein; 231 AA.
XX
XX  ADB65422;
XX
XX  04-DEC-2003 (first entry)
XX
XX  Human protein encoded by clone TEST120180600.
XX
XX  Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;
XX  cell regeneration; membrane protein; signal transduction-related protein;
XX  transcription-related protein; osteoporosis; neurological disease;
XX  cancer; tumour.
XX
XX  Homo sapiens.
XX
XX  EP1308459-A2.
XX
XX  07-MAY-2003.
XX
XX  28-MAR-2002; 2002EP-00007401.
XX
XX  05-NOV-2001; 2001JP-00379298.
XX
XX  25-JAN-2002; 2002US-00350978.
XX
XX  (HELI-) HELIX RES INST.
XX
XX  (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
XX  Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
XX

```

PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
 PI Seki N, Yoshikawa I, Otsuka M, Nagahari K, Masuho Y;
 XX WPI, 2003-450961/43.
 DR N-PSDB; ADB63452.
 XX
 PT New polynucleotides and polypeptides, useful for developing a diagnostic
 PT marker or medicines for regulation of their expression and activity, or
 PT as targets of gene therapy.
 XX
 PS Claim 1, Page; 222pp; English.

XX The invention discloses a polynucleotide comprising a sequence selected
 CC from 1970 fully defined nucleotide sequences which encode novel
 CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
 CC or its partial peptide, an antibody binding to the polypeptide or peptide
 CC of the polynucleotide, immunologically assaying the polypeptide or
 CC peptide of the polynucleotide by contacting the polypeptide or peptide
 CC with the antibody of the encoded protein, and observing the binding
 CC between the two, a transformant carrying the polynucleotide in an
 CC expressible manner and an antisense polynucleotide. The oligonucleotide
 CC is useful as a primer for synthesizing the polynucleotide, or as a probe
 CC for detecting the polynucleotide. The polynucleotide, or as a probe
 CC proteins are useful as pharmaceutical agents and many disease-related
 CC genes may be included in them, for developing a diagnostic marker or
 CC medicines for regulation of their expression and activity, or as targets
 CC of gene therapy. The genes are involved in tissue and/or cell
 CC regeneration. Membrane proteins, signal transduction-related proteins,
 CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
 CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
 CC the activity or expression of the encoded protein to treat diseases. The
 CC sequence presented is a protein of the invention. Note: Some of the
 CC sequence data for this patent is not represented in the printed
 CC specification, but is based on sequence information supplied by the
 CC European Patent Office.
 XX

SO Sequence 231 AA;

Query Match 85.7%; Score 36; DB 7; Length 231;
 Best Local Similarity 62.5%; Pred. No. 1.3e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKHPOPER 8
 DB 100 RHHPOPER 107

RESULT 7
 AAB75608
 ID AAB75608 standard; protein; 313 AA.
 XX
 AC AAB75608;
 XX
 DT 10-APR-2001 (first entry)
 XX
 DE Human cancer associated antigen precursor HOM-TES-85 SEQ ID NO:11.
 XX
 KW Human; cancer associated antigen precursor; cancer associated antigen;
 KW seminoma; HLA; human leukocyte antigen; cytostatic; gene therapy;
 KW vaccine; cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200100874-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-US017207.
 XX
 PR 30-JUN-1999; 99US-00346498.
 XX
 PA (LUDWIG) LUDWIG INST CANCER RES.

XX
 PI Sahin U, Tureci O, Pfreundschuh M;
 XX
 XX WPI, 2001-112465/12.
 DR N-PSDB; AAF26957.
 XX
 PT Diagnosing a disorder characterized by expression of a human cancer
 PT associated antigen precursor, comprises detecting interaction of an agent
 PT with a nucleic acid molecule encoding the antigen precursor.
 XX
 PS Example 2, Page 97; 126pp; English.

XX The present invention describes a method for diagnosing a disorder
 CC characterised by expression of a human cancer associated antigen (CAA)
 CC precursor (I) coded by a NA Group 1 nucleic acid molecule (NI) comprising
 CC contacting the biological sample with an agent (A) that specifically
 CC binds to NI, (I) or its fragment, complexed with an human leukocyte
 CC antigen (HLA) molecule and determining the interaction between the agent
 CC and NI or (I). (I) has cytostatic activity and can be used in gene
 CC therapy and vaccine production. The method can be used for treating a
 CC subject with a condition characterised by expression of (I) in cells of a
 CC subject. AAB75607 and AAB75608 represent proteins from human cancer
 CC associated antigen precursors, and AAB75609 to AAB75802 represent HLA
 CC class I binding motifs in human cancer associated antigen precursors
 XX given in the exemplification of the present invention

SO Sequence 313 AA;

Query Match 85.7%; Score 36; DB 4; Length 313;
 Best Local Similarity 62.5%; Pred. No. 1.8e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKHPOPER 8
 DB 182 RHHPOPER 189

RESULT 8
 ADS41924
 ID ADS41924 standard; protein; 359 AA.
 XX
 AC ADS41924;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Bacterial polypeptide #20354.
 XX
 KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide resistance;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polypeptide.
 XX
 OS Bacteria.
 XX
 PN US2003233675-A1.
 XX
 PD 18-DEC-2003.
 XX
 PF 20-FEB-2003; 2003US-00369493.
 XX
 PR 21-FEB-2002; 2002US-0360039P.
 XX
 PA (CAOY/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

DR WPI; 2004-061375/06.
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 20354; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 359 AA;
XX
Query Match 85.7%; Score 36; DB 8; Length 359;
Best Local Similarity 62.5%; Pred. No. 2.1e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 KKHPOFER 8
DB 311 RKHPEFER 318
XX
RESULT 9
ADN18635
ID ADN18635 standard; protein; 360 AA.
XX
AC ADN18635;
XX
DT 02-DEC-2004 (first entry)
XX
DE Bacterial polypeptide #1288.
XX
KW Recombinant DNA construct; transformed plant; improved plant property;
KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;
KW pathogen tolerance; pest tolerance; plant disease resistance;
KW cell cycle pathway modification; plant growth regulator;
KW homologous recombination; seed oil yield; protein yield; carbohydrate;
KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW bacterial polypeptide.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.

PA (HINK/) HINKLE G J.
PA (SLATER/) SLATER S C.
PA (CHEN/) CHEN X. S.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR WPI; 2004-061375/06.
XX
XX New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 1288; 122pp; English.
XX
CC The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 360 AA;
XX
Query Match 85.7%; Score 36; DB 8; Length 360;
Best Local Similarity 62.5%; Pred. No. 2.1e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 KKHPOFER 8
DB 311 RKHPEFER 318
XX
RESULT 10
ABM83512
ID ABM83512 standard; protein; 598 AA.
XX
AC ABM83512;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human diagnostic and therapeutic protein SEQ ID NO:3761.
XX
KW gene therapy; human diagnostic and therapeutic polynucleotide; ditnp.
XX
OS Homo sapiens.
XX
PN WO2004023973-A2.
XX
PD 25-MAR-2004.
XX
PF 12-SEP-2003; 2003WO-US028227.
XX
PR 12-SEP-2002; 2002US-0410259P.
XX
PA 12-SEP-2002; 2002US-0410260P.

XX (INCY-) INCYTE CORP.
 PA Schmidt JP, Wright RJ, Brune CM, Marjanovic MM, Shen F;
 PI Hatheshorne TA, Suchorolski MT, Altus CM, Plets SU, Elder LV;
 PI Mooney EM, Deleage AM, Panesar IS, Barville SC, Reddy TP;
 PI Stevens KA, Blanchard JL, Panzer SR, Wang X, Au AP, Gerslin EH;
 PI Peralta CH, Anderson SB, Ricou P, Shen EJ, Wu MC, Stuve IL;
 PI Lagace RE, Spiro PA, Stewart EA, Wingrove J, Vitt UA, Kitton ES,
 PI Xu Y, Kwong M, Policky JL, Hurwitz BL, Ma Y, Jackson JL, Gietzen D,
 PI Pacury S, Shi X, Suarez CJ;
 XX WPI: 2004-329368/30.
 DR N-PSDB; ACN42164.
 XX
 PT New diagnostic and therapeutic polynucleotides and polypeptides, useful
 PT in diagnosing a condition, disease or disorder associated with human
 PT molecules, e.g. autoimmune or inflammatory disorders, in gene therapy or
 PT in gene mapping.
 XX
 PS Claim 27; Page; 190pp; English.
 XX
 CC The invention relates to novel diagnostic and therapeutic polynucleotides
 CC selected from one of the 2722 sequences defined in the specification. A
 CC polynucleotide of the invention may have a use in gene therapy. The human
 CC diagnostic and therapeutic polynucleotides (dthp) or polypeptides may be
 CC used to diagnose a particular condition, disease or disorder associated
 CC with human molecules, e.g. cell proliferative disorders, or
 CC autoimmune/inflammatory disorder, developmental disorder, endocrine
 CC disorder, neurological disorders, gastrointestinal disorders, or
 CC infections caused by virus, bacteria, fungi or parasite. The dthp
 CC molecules may also be used in genetic mapping, in identifying individuals
 CC from minute biological samples, in detecting single nucleotide
 CC polymorphisms, as molecular weight markers, and for somatic or germline
 CC gene therapy. The present sequence represents a dthp protein of the
 CC invention. Note: The sequence data for this patent is not represented in
 CC the printed specification, but was obtained in electronic format directly
 CC from WFO at www.wipo.int/pct/en/sequences/listing.htm
 XX
 SQ Sequence 598 AA;
 XX
 Query Match 85.7%; Score 36; DB 0; Length 598;
 Best Local Similarity 62.5%; Pred. No. 3.7e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KXHPFER 8
 DB 417 KTHPOYEX 424
 XX
 RESULT 11
 ADN25797
 ID ADN25797 standard; protein; 378 AA.
 XX
 AC ADN25797;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Bacterial polypeptide #8450.
 XX
 KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polypeptide.
 XX
 OS Bacteria.
 XX
 PN US200333675-A1.
 XX
 PD 18-DEC-2003.

XX
 PF 20-FEB-2003; 2003US-00369493.
 XX
 XX 21-FEB-2002; 2002US-0360039P.
 XX
 PA (CAOY/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX
 PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 PI WPI: 2004-061375/06.
 DR
 XX
 PT New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 PT
 XX
 PS Claim 1; SEQ ID NO 8450; 122pp; English.
 XX
 CC The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polypeptide used in the
 CC scope of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification but was obtained in electronic
 CC format from USPTO at segdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 378 AA;
 XX
 Query Match 83.3%; Score 35; DB 8; Length 378;
 Best Local Similarity 75.0%; Pred. No. 3.4e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 KXHPFER 8
 DB 259 KQHPFER 266
 XX
 RESULT 12
 ADS23840
 ID ADS23840 standard; protein; 455 AA.
 XX
 AC ADS23840;
 XX
 DT 02-DEC-2004 (first entry)
 XX
 DE Bacterial polypeptide #12873.
 XX
 KW Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;

KW bacterial polypeptide.
XX
OS Bacteria.
XX
PN US2003233675-A1.
XX
PD 18-DEC-2003.
XX
PF 20-FEB-2003; 2003US-00369493.
XX
PR 21-FEB-2002; 2002US-0360039P.
XX
PA (CAOY/) CAO Y.
XX (HINK/) HINKLE G J.
PA (SLAT/) SLATER S C.
PA (CHEN/) CHEN X.
PA (GOLD/) GOLDMAN B S.
XX
PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR WPI; 2004-061375/06.
XX
PT New recombinant DNA construct comprising a promoter positioned to provide
PT for expression of a polynucleotide encoding a polypeptide from a
PT microbial source, useful for producing plants with improved properties.
XX
PS Claim 1; SEQ ID NO 12873; 122pp; English.
XX
XX The invention relates to a recombinant DNA construct comprising a
CC promoter functional in a plant cell, where the promoter is positioned to
CC provide for expression of a polynucleotide encoding a polypeptide from a
CC microbial source. The invention also relates to a transformed plant
CC comprising the recombinant DNA construct and a method of producing a
CC transformed plant having an improved property. The plant is a crop plant
CC such as maize or soybean. The method of producing a transformed plant
CC having an improved property comprises transforming a plant with the
CC recombinant DNA construct and growing the transformed plant, where the
CC polynucleotide or polypeptide is useful for improving plant properties.
CC The recombinant DNA construct is useful for producing plants with
CC improved plant properties, e.g. improved cold, heat or drought tolerance,
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC increased resistance to plant disease, better growth rate by modification
CC of the cell cycle pathway with plant growth regulators, increased rate of
CC homologous recombination, modified seed oil or protein yield and/or
CC content, improved yield by modification of carbohydrate, nitrogen or
CC phosphorus use and/or uptake, by modification of photosynthesis or by
CC providing improved plant growth and development under at least one stress
CC condition, improved lignin production or improved galactomannan
CC production. This sequence represents a bacterial polypeptide used in the
CC scope of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 455 AA;
XX
Query Match 83.3%; Score 35; DB 8; Length 455;
Best Local Similarity 62.5%; Pred. No. 4.2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
OY 1 KKHPOFER 8
DB 277 KKHPRFR 284
XX
RESULT 13
ID AAM59212
XX AAM59212 standard; peptide; 8 AA.
AC AAM59212;
XX
XX 27-AUG-1998 (first entry)
DT
XX Streptavidin tagged peptide ligand #2.
DE

XX
XX Streptavidin; ligand; binding affinity; mutant; isolation; purification;
KW recover; immobilise.
XX
OS Synthetic.
XX
PN EP835934-A2.
XX
PD 15-APR-1998.
XX
PF 09-OCT-1997; 97EP-00117504.
XX
PR 10-OCT-1996; 96DE-01041876.
XX
PA (BIOA-) INST BIOANALYTIK GMBH.
XX
PI Skerra A, Voess S;
XX
DR WPI; 1998-218868/20.
XX
PT Streptavidin mutants with higher binding affinity for peptide ligands -
PT have mutation in amino acid region 44-53, used to isolate, purify or
PT determine fusion proteins including these ligands.
XX
PS Claim 10; Page 11; 21pp; German.
XX
XX AAM59211 and AAM59212 are ligands used in a method to assay binding
CC affinity of streptavidin mutants. These mutants have a mutation within
CC the amino acid (aa) region 44-53 of the wild-type protein show a higher
CC binding affinity than the wild-type for peptide ligands that include the
CC sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z
CC are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin
CC mutants can be used to isolate, purify and determine proteins or to
CC determine/recover substances that contain streptavidin-binding groups.
CC Such compounds may also be used to immobilise fusions on microtitre
CC plates, microbeads or sensor chips
XX
SQ Sequence 8 AA;
XX
Query Match 81.0%; Score 34; DB 2; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.8e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
OY 2 KKHPOFER 8
DB 2 SHPOFEK 8
XX
RESULT 14
ID AAB35433
XX AAB35433 standard; peptide; 8 AA.
AC AAB35433;
XX
DT 23-MAY-2001 (first entry)
XX
DE Epitope peptide #3.
XX
KW Nascent protein detection; protein analysis; aminoacylated tRNA;
KW BODIPY marker; disease diagnosis.
XX
OS Unidentified.
XX
PN WO200114578-A1.
XX
PD 01-MAR-2001.
XX
PF 23-AUG-2000; 2000WO-US023233.
XX
XX 25-AUG-1999; 99US-00382736.
PR 25-AUG-1999; 99US-00382950.
XX
XX (AMBE-) AMBERGEN INC.
PA

```

XX  Rothschld KJ, Gite S, Olejnik J;
PI
XX  WPI; 2001-168972/17.
DR
XX  Method for detecting nascent proteins by fluorescence comprises
PT  misaminoacylating a tRNA molecule with a marker compound, useful for
PT  detecting mutations in proteins, e.g. cancer.
XX
PS  Disclosure; Page 47; 204pp; English.
XX
CC  The present invention describes a method of detecting nascent proteins
CC  involving aminoacylating a tRNA molecule with a 4,4-difluoro-4-bora-3a,4a
CC  -diaz-a-s-indacene (BODIPY) marker leading to the production of a
CC  misaminoacylated tRNA. This enables the detection, isolation and analysis
CC  of nascent proteins using UV without the usual accompanying radioactivity
CC  problems. It may be used to detect mutations, for example in cancer.
CC  Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer
XX
SQ  Sequence 8 AA;

```

```

Query Match      81.0%; Score 34; DB 4; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.8e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

OY  2 XHPQFER 8
    :|||||:
DB  2 SHPQPER 8

```

```

RESULT 15
AAB35442
ID  AAB35442 standard; peptide; 8 AA.
XX
AC  AAB35442;
XX
DT  23-MAY-2001 (first entry)
XX
DE
XX
DE  Nascent protein detection method related peptide #4.
XX
KM  Nascent protein detection; protein analysis; aminoacylated tRNA;
XX  BODIPY marker; disease diagnosis.
XX
OS  Unidentified.
XX
PN  MO280114578-A1.
XX
PD  01-MAR-2001.
XX
PF  23-AUG-2000; 2000WO-US023233.
XX
PR  25-AUG-1999; 99US-00382736.
XX  25-AUG-1999; 99US-00382950.
XX
PA  (AMBE-) AMBERGEN INC.
XX
PI  Rothschld KJ, Gite S, Olejnik J;
XX
DR  WPI; 2001-168972/17.
XX

```

```

PT  Method for detecting nascent proteins by fluorescence comprises
PT  misaminoacylating a tRNA molecule with a marker compound, useful for
PT  detecting mutations in proteins, e.g. cancer.
XX
PS  Example 22; Page 153; 204pp; English.
XX

```

```

CC  The present invention describes a method of detecting nascent proteins
CC  involving aminoacylating a tRNA molecule with a 4,4-difluoro-4-bora-3a,4a
CC  -diaz-a-s-indacene (BODIPY) marker leading to the production of a
CC  misaminoacylated tRNA. This enables the detection, isolation and analysis
CC  of nascent proteins using UV without the usual accompanying radioactivity
CC  problems. It may be used to detect mutations, for example in cancer,
CC  Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer
CC

```

```

XX
SQ  Sequence 8 AA;

```

```

Query Match      81.0%; Score 34; DB 4; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.8e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

OY  2 XHPQFER 8
    :|||||:
DB  2 SHPQPER 8

```

```

Search completed: March 2, 2005, 13:02:45
Job time : 37.0325 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 24.878 Seconds
(without alignments)
105.489 Million cell updates/sec

Title: SEQ6
Perfect score: 42
Sequence: 1 kxhpqfer 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	85.7	231	US-10-104-047-3576	Sequence 3576, Ap
2	36	85.7	359	US-10-369-493-20354	Sequence 20354, A
3	36	85.7	360	US-10-369-493-1288	Sequence 1288, Ap
4	35	83.3	36	US-10-026-578B-11	Sequence 11, Appl
5	35	83.3	378	US-10-369-493-8450	Sequence 8450, Ap
6	35	83.3	455	US-10-369-493-12873	Sequence 12873, A
7	34	81.0	8	US-09-809-517A-9	Sequence 9, Appl
8	34	81.0	9	US-09-873-145-7	Sequence 7, Appl
9	34	81.0	8	US-09-813-197-8	Sequence 8, Appl
10	34	81.0	8	US-10-104-218-5	Sequence 5, Appl
11	34	81.0	8	US-10-208-357-9	Sequence 9, Appl
12	34	81.0	8	US-10-001-934-8	Sequence 8, Appl
13	34	81.0	8	US-10-026-578B-2	Sequence 2, Appl

14	34	81.0	8	US-10-026-578B-9	Sequence 9, Appl
15	34	81.0	8	US-10-174-368A-7	Sequence 7, Appl
16	34	81.0	8	US-10-264-177-8	Sequence 8, Appl
17	34	81.0	8	US-10-339-712-8	Sequence 8, Appl
18	34	81.0	8	US-10-339-712-67	Sequence 67, Appl
19	34	81.0	8	US-10-275-046-4	Sequence 4, Appl
20	34	81.0	8	US-10-425-000-76	Sequence 76, Appl
21	34	81.0	8	US-10-425-999-25	Sequence 25, Appl
22	34	81.0	8	US-10-358-283-23	Sequence 23, Appl
23	34	81.0	8	US-10-628-432-41	Sequence 41, Appl
24	34	81.0	8	US-10-494-248-17	Sequence 17, Appl
25	34	81.0	8	US-10-634-645-11	Sequence 11, Appl
26	34	81.0	8	US-10-719-523-8	Sequence 8, Appl
27	34	81.0	9	US-09-983-067-3	Sequence 3, Appl
28	34	81.0	10	US-09-809-517A-6	Sequence 6, Appl
29	34	81.0	10	US-10-147-211A-20	Sequence 20, Appl
30	34	81.0	11	US-10-354-983-29	Sequence 29, Appl
31	34	81.0	11	US-10-628-432-25	Sequence 25, Appl
32	34	81.0	13	US-10-338-592-30	Sequence 30, Appl
33	34	81.0	21	US-09-809-517A-30	Sequence 30, Appl
34	34	81.0	22	US-09-809-517A-33	Sequence 33, Appl
35	34	81.0	24	US-09-809-517A-31	Sequence 31, Appl
36	34	81.0	24	US-10-026-578B-3	Sequence 3, Appl
37	34	81.0	24	US-10-026-578B-4	Sequence 4, Appl
38	34	81.0	25	US-09-809-517A-34	Sequence 34, Appl
39	34	81.0	36	US-10-026-578B-10	Sequence 10, Appl
40	34	81.0	117	US-09-977-137A-4	Sequence 4, Appl
41	34	81.0	117	US-09-977-137A-5	Sequence 5, Appl
42	34	81.0	117	US-09-977-137A-7	Sequence 7, Appl
43	34	81.0	117	US-09-977-137A-8	Sequence 8, Appl
44	34	81.0	117	US-09-977-137A-9	Sequence 9, Appl
45	34	81.0	117	US-09-977-137A-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-10-104-047-3576
; Sequence 3576, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1e1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT FILING DATE: US/10/104,047
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: Patencin Ver. 2.1
; SEQ ID NO 3576
; LENGTH: 231
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3576

Query Match 85.7%; Score 36; DB 15; Length 231;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KXHPQFER 8
DB 100 RHHPOYER 107

RESULT 2
US-10-369-493-20354
; Sequence 20354, Application US/10369493
; Publication No. US20030236375A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.

APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 20354
 LENGTH: 359
 TYPE: PRT
 ORGANISM: *Pyrococcus horikoshii*
 US-10-369-493-20354

Query Match 85.7%; Score 36; DB 15; Length 359;
 Best Local Similarity 62.5%; Pred. No. 2.5e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKHPOFER 8
 DB 311 RKHPFER 318

RESULT 3
 US-10-369-493-1288
 Sequence 1288, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 1288
 LENGTH: 360
 TYPE: PRT
 ORGANISM: *Pyrococcus horikoshii*
 US-10-369-493-1288

Query Match 85.7%; Score 36; DB 15; Length 360;
 Best Local Similarity 62.5%; Pred. No. 2.5e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKHPOFER 8
 DB 311 RKHPFER 318

RESULT 4
 US-10-026-578B-11
 Sequence 11, Application US/10026578B
 Publication No. US20030083474A1
 GENERAL INFORMATION:
 APPLICANT: IBA (GmbH)
 APPLICANT: Schmidt, Thomas
 TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
 FILE REFERENCE: 100810.01u81
 CURRENT APPLICATION NUMBER: US/10/026,578B
 CURRENT FILING DATE: 2002-11-11
 PRIOR APPLICATION NUMBER: DE 101 13 776.1
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: PCT/EP01/11846

PRIOR FILING DATE: 2001-10-12
 NUMBER OF SEQ ID NOS: 14
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 11
 LENGTH: 36
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Synthetic Peptide
 NAME/KEY: MISC FEATURE
 LOCATION: (9)..(28)
 OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more
 OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will
 FEATURE:
 NAME/KEY: misc feature
 OTHER INFORMATION: Artificial Sequence represents peptide binding module
 US-10-026-578B-11

Query Match 83.3%; Score 35; DB 14; Length 36;
 Best Local Similarity 62.5%; Pred. No. 38;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 KKHPOFER 8
 DB 29 QSHPOFER 36

RESULT 5
 US-10-369-493-8450
 Sequence 8450, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 CURRENT FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 PRIOR FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 8450
 LENGTH: 378
 TYPE: PRT
 ORGANISM: *Ralstonia metallidurans*
 US-10-369-493-8450

Query Match 83.3%; Score 35; DB 15; Length 378;
 Best Local Similarity 75.0%; Pred. No. 3.9e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKHPOFER 8
 DB 259 KKHPOFER 266

RESULT 6
 US-10-369-493-12873
 Sequence 12873, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

```

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 12873
; LENGTH: 455
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(455)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-12873

Query Match      83.3%; Score 35; DB 15; Length 455;
Best Local Similarity 62.5%; Pred. No. 4.7e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY      1 XHPOFER 8
DB      277 KHPRFOR 284

RESULT 7
US-09-809-517A-9
; Sequence 9, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote
; FILE REFERENCE: MOREHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-9

Query Match      81.0%; Score 34; DB 9; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPOFER 8
DB      2 SHPOFER 8

RESULT 8
US-09-973-145-7
; Sequence 7, Application US/09973145
; Patent No. US20020132248A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
; FILE REFERENCE: AMBR-06819
; CURRENT APPLICATION NUMBER: US/09/973,145
; CURRENT FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 09/382,950

```

```

; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Synthetic
US-09-973-145-7

Query Match      81.0%; Score 34; DB 9; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPOFER 8
DB      2 SHPOFER 8

RESULT 9
US-09-813-197-8
; Sequence 8, Application US/09813197
; Publication No. US2005009013A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot
; FILE REFERENCE: AMBR-03951
; CURRENT APPLICATION NUMBER: US/09/813,197
; CURRENT FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-813-197-8

Query Match      81.0%; Score 34; DB 12; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPOFER 8
DB      2 SHPOFER 8

RESULT 10
US-10-104-218-5
; Sequence 5, Application US/10104218
; Publication No. US20020177196A1
; GENERAL INFORMATION:
; APPLICANT: MAIER, Thomas
; APPLICANT: GAEBERT, Carsten
; TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
; FILE REFERENCE: MAIER, T. ET AL.-2
; CURRENT APPLICATION NUMBER: US/10/104,218
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 16 881.0
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 21 515.0
; PRIOR FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 8

```

TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: PEPTIDE
LOCATION: (1)-(8)
OTHER INFORMATION: Streptagrit affinity peptide for protein purification
US-10-104-218-5

Query Match 81.0%; Score 34; DB 13; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8
:|||||:
DB 2 SHPOFER 8

RESULT 11
US-10-208-357-9
Sequence 9, Application US/10208357
Publication No. US20020182687A1
GENERAL INFORMATION:
APPLICANT: Kurz, Markus
APPLICANT: Lohse, Peter
APPLICANT: Wagner, Richard
TITLE OF INVENTION: Peptide Acceptor Ligation Methods
FILE REFERENCE: 50036/031002
CURRENT APPLICATION NUMBER: US/10/208,357
CURRENT FILING DATE: 2002-07-30
PRIOR APPLICATION NUMBER: US/09/619,103
PRIOR FILING DATE: 2000-07-19
PRIOR FILING DATE: 1999-07-27
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: designed sequence to act as an identifying tag
US-10-208-357-9

Query Match 81.0%; Score 34; DB 13; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8
:|||||:
DB 2 SHPOFER 8

RESULT 12
US-10-001-934-8
Sequence 8, Application US/10001934
Publication No. US20030032782A1
GENERAL INFORMATION:
APPLICANT: NAGY, ZOLTAN
TITLE OF INVENTION: HUMAN PEPTIDES/PROTEINS CAUSING OR LEADING TO THE
KILLING OF CELLS INCLUDING LYMPHOID TUMOR CELLS
FILE REFERENCE: GPCG-P01-003
CURRENT APPLICATION NUMBER: US/10/001,934
CURRENT FILING DATE: 2001-11-15
NUMBER OF SEQ ID NOS: 63
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 8
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-001-934-8

Query Match 81.0%; Score 34; DB 14; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8
:|||||:
DB 2 SHPOFER 8

RESULT 13
US-10-026-578B-2
Sequence 2, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
APPLICANT: Schmidt, Thomas
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
CURRENT FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
PRIOR FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 8
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-2

Query Match 81.0%; Score 34; DB 14; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8
:|||||:
DB 2 SHPOFER 8

RESULT 14
US-10-026-578B-9
Sequence 9, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
APPLICANT: Schmidt, Thomas
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
CURRENT FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
PRIOR FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 9
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-9

US-10-026-578B-9

Query Match 81.0%; Score 34; DB 14; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPOFER 8
:|||||:
Db 2 SHPOFEK 8

RESULT 15

US-10-174-368A-7
; Sequence 7, Application US/10174368A
; Publication No. US20030092031A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Method for the Detection, Analysis and Isolation of Nascent Pro
; FILE REFERENCE: AMBR-07145
; CURRENT APPLICATION NUMBER: US/10/174,368A
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: 10/049,332
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/23233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patencin version 3.1
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Synthetic
US-10-174-368A-7

Query Match 81.0%; Score 34; DB 14; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPOFER 8
:|||||:
Db 2 SHPOFEK 8

Search completed: March 2, 2005, 14:18:50
Job time : 24.878 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 9.10569 Seconds
(without alignments)
65.585 Million cell updates/sec

Title: SEQ6

Perfect score: 42

Sequence: 1 kxnpqfer 8

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*\n2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*\n3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*\n4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*\n5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*\n6: /cgn2_6/prodata/1/1aa/backfill1.pep:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	482	4 US-09-107-532A-4868	Sequence 4868, Ap
2	34	81.0	8	3 US-08-948-097-2	Sequence 2, Appl
3	34	81.0	8	3 US-09-382-950-7	Sequence 7, Appl
4	34	81.0	8	3 US-09-382-736B-8	Sequence 8, Appl
5	34	81.0	8	4 US-09-619-103-9	Sequence 9, Appl
6	34	81.0	8	4 US-10-104-218-5	Sequence 5, Appl
7	34	81.0	8	4 US-09-809-517A-9	Sequence 6, Appl
8	34	81.0	10	4 US-09-809-517A-6	Sequence 9, Appl
9	34	81.0	13	2 US-08-687-559-30	Sequence 30, Appl
10	34	81.0	13	4 US-09-401-415-30	Sequence 30, Appl
11	34	81.0	21	4 US-09-809-517A-30	Sequence 30, Appl
12	34	81.0	22	4 US-09-809-517A-33	Sequence 33, Appl
13	34	81.0	24	4 US-09-809-517A-31	Sequence 31, Appl
14	34	81.0	25	4 US-09-809-517A-34	Sequence 34, Appl
15	34	81.0	117	4 US-09-977-137A-4	Sequence 4, Appl
16	34	81.0	117	4 US-09-977-137A-5	Sequence 5, Appl
17	34	81.0	117	4 US-09-977-137A-7	Sequence 7, Appl
18	34	81.0	117	4 US-09-977-137A-8	Sequence 8, Appl
19	34	81.0	117	4 US-09-977-137A-9	Sequence 9, Appl
20	34	81.0	117	4 US-09-977-137A-10	Sequence 10, Appl
21	34	81.0	117	4 US-09-977-137A-11	Sequence 11, Appl
22	34	81.0	117	4 US-09-977-137A-12	Sequence 12, Appl
23	34	81.0	118	4 US-09-977-137A-6	Sequence 6, Appl
24	34	81.0	250	4 US-09-803-286A-7	Sequence 7, Appl
25	34	81.0	337	4 US-09-665-479A-8	Sequence 8, Appl
26	34	81.0	399	3 US-09-553-498-2	Sequence 2, Appl
27	34	81.0	399	4 US-09-618-869-2	Sequence 2, Appl

28	34	81.0	410	4 US-09-248-796A-18368	Sequence 18368, A
29	34	81.0	428	3 US-09-052-778-12	Sequence 12, Appl
30	34	81.0	455	2 US-08-588-983-7	Sequence 7, Appl
31	34	81.0	455	2 US-08-588-976-7	Sequence 7, Appl
32	34	81.0	918	2 US-08-588-983-14	Sequence 14, Appl
33	34	81.0	918	2 US-08-588-976-14	Sequence 14, Appl
34	34	81.0	919	2 US-08-588-983-9	Sequence 9, Appl
35	34	81.0	919	2 US-08-588-983-12	Sequence 12, Appl
36	34	81.0	919	2 US-08-588-976-9	Sequence 9, Appl
37	34	81.0	919	2 US-08-588-976-12	Sequence 12, Appl
38	34	81.0	1220	3 US-08-930-996A-2	Sequence 2, Appl
39	33	78.6	114	4 US-09-270-767-42719	Sequence 42719, A
40	33	78.6	209	4 US-09-583-110-2388	Sequence 2388, Ap
41	33	78.6	209	4 US-09-107-433-3989	Sequence 3989, Ap
42	33	78.6	276	3 US-08-961-083-134	Sequence 134, App
43	33	78.6	276	3 US-09-536-784-134	Sequence 134, App
44	33	78.6	462	4 US-09-248-796A-18226	Sequence 18226, A
45	33	78.6	741	4 US-09-252-991A-20098	Sequence 20098, A

ALIGNMENTS

RESULT 1
US-09-107-532A-4868
Sequence 4868, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-8277
TELEFAX: (781)893-5007
INFORMATION FOR SEQ ID NO: 4868:
SEQUENCE CHARACTERISTICS:
LENGTH: 482 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...482
SEQUENCE DESCRIPTION: SEQ ID NO: 4868:
US-09-107-532A-4868

Query Match
Best Local Similarity 92.9%; Score 39; DB 4; Length 482;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 XHPQFER 8
DB 409 KKHPEFER 416

RESULT 2
US-08-948-097-2
Sequence 2, Application US/08948097C
Patent No. 6103493
GENERAL INFORMATION:
APPLICANT: Skerra, Arne
APPLICANT: Voss, Selma
TITLE OF INVENTION: Streptavidin Mutelins
FILE REFERENCE: HUBR 1119
CURRENT APPLICATION NUMBER: US/08/948,097C
CURRENT FILING DATE: 1997-10-09
EARLIER APPLICATION NUMBER: DE 196 41 876.3
NUMBER OF SEQ ID NOS: 17
NUMBER OF SEQ ID NOS: 17
SEQ ID NO 2
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial sequence
FEATURE:
NAME/KEY: BINDING
OTHER INFORMATION: Binding ligand for streptavidin
US-08-948-097-2

Query Match
Best Local Similarity 81.0%; Score 34; DB 3; Length 8;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFER 8
DB 2 SHPQFER 8

RESULT 3
US-09-382-950-7
Sequence 7, Application US/09382950
Patent No. 6303337
GENERAL INFORMATION:
APPLICANT: Rothechild, Kenneth
APPLICANT: Gite, Sadanand
APPLICANT: Olejnik, Jerzy
TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
FILE REFERENCE: AMBER-03879
CURRENT APPLICATION NUMBER: US/09/382,950
CURRENT FILING DATE: 1999-08-25
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin version 3.0
SEQ ID NO 7
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial/Unknown
FEATURE:
NAME/KEY: misc feature
LOCATION: (..T)
OTHER INFORMATION: Synthetic
US-09-382-950-7

Query Match
Best Local Similarity 81.0%; Score 34; DB 3; Length 8;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFER 8
DB 2 SHPQFER 8

RESULT 4
US-09-382-736B-8
Sequence 8, Application US/09382736B
Patent No. 6306628
GENERAL INFORMATION:
APPLICANT: Rothechild, Kenneth
APPLICANT: Gite, Sadanand
APPLICANT: Olejnik, Jerzy
TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot
FILE REFERENCE: AMBER-03951
CURRENT APPLICATION NUMBER: US/09/382,736B
CURRENT FILING DATE: 1999-08-25
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin version 3.0
SEQ ID NO 8
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-09-382-736B-8

Query Match
Best Local Similarity 81.0%; Score 34; DB 3; Length 8;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFER 8
DB 2 SHPQFER 8

RESULT 5
US-09-619-103-9
Sequence 9, Application US/09619103
Patent No. 6429300
GENERAL INFORMATION:
APPLICANT: Kurz, Markus
APPLICANT: Lohse, Peter
APPLICANT: Wagner, Richard
TITLE OF INVENTION: Peptide Acceptor Ligation Methods
FILE REFERENCE: 50036/031002
CURRENT APPLICATION NUMBER: US/09/619,103
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 60/145,834
PRIOR FILING DATE: 1999-07-27
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 8
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: designed sequence to act as an identifying tag
US-09-619-103-9

Query Match
Best Local Similarity 81.0%; Score 34; DB 4; Length 8;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFER 8
DB 2 SHPQFER 8

RESULT 6
US-10-104-218-5
Sequence 5, Application US/10104218
Patent No. 6579705
GENERAL INFORMATION:
APPLICANT: MAIER, Thomas
APPLICANT: GABBERT, Carsten

```
/ TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
/ FILE REFERENCE: MAIER, T. ET AL.-2
/ CURRENT APPLICATION NUMBER: US/10/104,218
/ PRIOR FILING DATE: 2002-03-22
/ PRIOR APPLICATION NUMBER: German No. 6579705 101 16 881.0
/ PRIOR FILING DATE: 2001-04-04
/ PRIOR APPLICATION NUMBER: German No. 6579705 101 21 515.0
/ PRIOR FILING DATE: 2001-05-03
/ NUMBER OF SEQ ID NOS: 5
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 5
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ NAME/KEY: PEPTIDE
/ LOCATION: (1)..(8)
/ OTHER INFORMATION: StreptagII affinity peptide for protein purification
US-10-104-218-5
```

```
Query Match      81.0%; Score 34; DB 4; Length 8;
Best Local Similarity 71.4%; Pred. No. 4.1e+05;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHP0FER 8
        :|||||:
Db       2 SHPOFEK 8
```

```
RESULT 7
US-09-809-517A-9
/ Sequence 9, Application US/09809517A
/ Patent No. 6753136
/ GENERAL INFORMATION:
```

```
/ APPLICANT: Lohning, Corinna
/ TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on H
/ FILE REFERENCE: MORPHO/11
/ CURRENT APPLICATION NUMBER: US/09/809,517A
/ PRIOR FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: EP 99114072.4
/ PRIOR FILING DATE: 1999-07-20
/ PRIOR APPLICATION NUMBER: EP 00103551.8
/ PRIOR FILING DATE: 2000-02-18
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 9
/ LENGTH: 8
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-9
```

```
Query Match      81.0%; Score 34; DB 4; Length 8;
Best Local Similarity 71.4%; Pred. No. 4.1e+05;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHP0FER 8
        :|||||:
Db       2 SHPOFEK 8
```

```
RESULT 8
US-09-809-517A-6
/ Sequence 6, Application US/09809517A
/ Patent No. 6753136
/ GENERAL INFORMATION:
/ APPLICANT: Lohning, Corinna
/ TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on H
/ FILE REFERENCE: MORPHO/11
/ CURRENT APPLICATION NUMBER: US/09/809,517A
```

```
/ CURRENT FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: EP 99114072.4
/ PRIOR FILING DATE: 1999-07-20
/ PRIOR APPLICATION NUMBER: EP 00103551.8
/ PRIOR FILING DATE: 2000-02-18
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 6
/ LENGTH: 10
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-6
```

```
Query Match      81.0%; Score 34; DB 4; Length 10;
Best Local Similarity 71.4%; Pred. No. 1.6;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHP0FER 8
        :|||||:
Db       4 SHPOFEK 10
```

```
RESULT 9
US-08-687-559-30
/ Sequence 30, Application US/08687559
/ Patent No. 5955647
/ GENERAL INFORMATION:
```

```
/ APPLICANT: Fitcher, John H.
/ APPLICANT: Beachy, Roger N.
/ TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO
/ NUMBER OF SEQUENCES: 30
/ CORRESPONDENCE ADDRESSES:
/ ADDRESSEE: Fish & Richardson, P.C.
/ STREET: 4225 Executive Square, Suite 1400
/ CITY: La Jolla
/ STATE: CA
/ COUNTRY: US
/ ZIP: 92037
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: Windows95
/ SOFTWARE: FastSeq for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/687,559
/ FILING DATE: No. 5955647ember 18, 1996
/ CLASSIFICATION: 424
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US95/01467
/ FILING DATE: 03-FEB-1995
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Learn, June M.
/ REGISTRATION NUMBER: 31,238
/ REFERENCE/DOCKET NUMBER: 07302/011001
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 619-678-5099
/ TELEFAX: 619-678-5070
/ INFORMATION FOR SEQ ID NO: 30:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 13 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
US-08-687-559-30
```

```
Query Match      81.0%; Score 34; DB 2; Length 13;
Best Local Similarity 71.4%; Pred. No. 2.1;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHP0FER 8
```

Db :||||:|
7 GHPOFOR 13

```

RESULT 10
US-09-401-415-30
; Sequence 30, Application US/09401415
; Patent No. 6503732
; GENERAL INFORMATION:
; APPLICANT: THE SCRIPPS RESEARCH INSTITUTE
; TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO
; OVERPRODUCE PEPTIDES AND PROTEINS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Spensley Horn Jubas & Lubitz
; STREET: 1880 Century Park East, Suite 500
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/401,415
; FILING DATE: 21-Sep-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01467
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Boslich, June M.
; REGISTRATION NUMBER: 31,238
; REFERENCE/DOCKET NUMBER: PD-4074
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 13 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..13
; SEQUENCE DESCRIPTION: SEQ ID NO: 30:
US-09-401-415-30

Query Match      81.0%; Score 34; DB 4; Length 13;
Best Local Similarity 71.4%; Pred. No. 2.1;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8
Db 7 GHPOFOR 13

RESULT 11
US-09-809-517A-30
; Sequence 30, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0

```

```

; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO: 30
; LENGTH: 21
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30

Query Match      81.0%; Score 34; DB 4; Length 21;
Best Local Similarity 71.4%; Pred. No. 3.5;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8
Db 15 SHPOFER 21

RESULT 12
US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO: 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

Query Match      81.0%; Score 34; DB 4; Length 22;
Best Local Similarity 71.4%; Pred. No. 3.7;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8
Db 16 SHPOFER 22

RESULT 13
US-09-809-517A-31
; Sequence 31, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0

```

```
/ SEQ ID NO 31
/ LENGTH: 24
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31
```

```
Query Match      81.0%; Score 34; DB 4; Length 24;
Best Local Similarity 71.4%; Pred. No. 4.1;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      2 XHPQFER 8
        :|||||:
Db      18 SHPOFEK 24
```

```
RESULT 14
US-09-809-517A-34
/ Sequence 34, Application US/09809517A
/ Patent No. 6753136
/ GENERAL INFORMATION:
/ APPLICANT: Lohming, Corinna
/ TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on H
/ FILE REFERENCE: MORPHO/11
/ CURRENT APPLICATION NUMBER: US/09/809,517A
/ CURRENT FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: EP 99114072.4
/ PRIOR FILING DATE: 1999-07-20
/ PRIOR APPLICATION NUMBER: EP 00103551.8
/ PRIOR FILING DATE: 2000-02-18
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 34
/ LENGTH: 25
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34
```

```
Query Match      81.0%; Score 34; DB 4; Length 25;
Best Local Similarity 71.4%; Pred. No. 4.3;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      2 XHPQFER 8
        :|||||:
Db      19 SHPOFEK 25
```

```
RESULT 15
US-09-977-137A-4
/ Sequence 4, Application US/09977137A
/ Patent No. 6750042
/ GENERAL INFORMATION:
/ APPLICANT: Summers, Anne O.
/ APPLICANT: Casagiat, Jonathan
/ TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 4
/ LENGTH: 117
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
```

US-09-977-137A-4

```
Query Match      81.0%; Score 34; DB 4; Length 117;
Best Local Similarity 71.4%; Pred. No. 23;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      2 XHPQFER 8
        :|||||:
Db      111 SHPOFEK 117
```

Search completed: March 2, 2005, 12:25:34
Job time : 10.1057 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 6.11382 Seconds
(without alignments)
125.901 Million cell updates/sec

Title: SEQ7

Perfect score: 44

Sequence: 1 rxnpgfsg 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR 79:*

1: p1r1:*

2: p1r2:*

3: p1r3:*

4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	88.6	131	2 S66537	Ig heavy chain V r
2	39	88.6	224	2 AF3382	alpha/beta hydrola
3	39	88.6	225	2 B97580	hypothetical prote
4	39	88.6	225	2 AH2800	conserved hypothet
5	39	88.6	379	2 S55900	DNAJ-like protein
6	39	88.6	379	2 T41633	p51 protein - fibro
7	39	88.6	436	1 S08156	gene T protein - m
8	39	88.6	512	2 H86216	protein T23G18.16
9	37	84.1	399	2 B69256	conserved hypothet
10	36	81.8	309	2 B87112	conserved hypothet
11	36	81.8	535	2 B82358	alkaline serine pr
12	35	79.5	222	2 S18106	type II site-speci
13	35	79.5	288	2 T12462	hypothetical prote
14	35	79.5	359	2 G84105	transcriptional regu
15	35	79.5	417	2 T01616	hypothetical prote
16	35	79.5	421	2 C83040	hypothetical prote
17	35	79.5	473	2 G72205	hypothetical prote
18	35	79.5	563	2 S70684	pyruvate decarboxy
19	35	79.5	564	1 S36363	pyruvate decarboxy
20	35	79.5	564	1 S50700	pyruvate decarboxy
21	35	79.5	857	2 AC2132	ferichrome-iron r
22	35	79.5	1061	2 A12579	conserved hypothet
23	35	79.5	1061	2 H97361	hypothetical prote
24	35	79.5	10797	2 T30192	probable peptide s
25	34	77.3	99	2 G84272	50S ribosomal prot
26	34	77.3	181	2 T07661	maturation protein
27	34	77.3	186	2 B86178	hypothetical prote
28	34	77.3	235	2 A96575	hypothetical prote
29	34	77.3	295	2 AG3129	transcription regu

30	34	77.3	295	2 B98158	appt protein (AF23
31	34	77.3	358	2 T48090	hypothetical prote
32	34	77.3	404	2 G75152	hypothetical prote
33	34	77.3	404	2 H71193	hypothetical prote
34	34	77.3	433	2 AD3570	galactarate dehydr
35	34	77.3	452	2 F83587	inner membrane pro
36	34	77.3	510	2 C69753	conserved hypothet
37	34	77.3	524	2 B97697	probable D-galacta
38	34	77.3	524	2 AE2923	D-galactarate dehy
39	33	75.0	95	1 R5H311	ribosomal protein
40	33	75.0	141	2 S08523	hypothetical prote
41	33	75.0	260	2 AB2080	hypothetical prote
42	33	75.0	296	2 B71369	probable sugar ABC
43	33	75.0	299	2 G82222	conserved hypothet
44	33	75.0	302	2 F86849	transcription regu
45	33	75.0	318	2 AG1020	cytochrome c-cye

ALIGNMENTS

RESULT 1

S66537 Ig heavy chain V region - mouse (fragment)

C/Species: Mus musculus (house mouse)

C/Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 20-Jun-2000

C/Accession: S66537

R/Tsictis, G.; Haase, W.; Engel, A.; Michel, H.

Eur. J. Biochem. 231, 823-830, 1995

A/Title: Isolation and structural characterization of trimeric cyanobacterial photosyst.

A/Reference number: S66536; PMID:95377318; PMID:7649183

A/Accession: S66537

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-131 <TS1>

C/Cross-references: EMBL:X88902; NID:9895869; PIDN:CAA61364.1; PID:g1103701

C/Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004

C/Accession: AF3382

R/Delvecchio, V.G.; Kaparatel, V.; Redkar, R.J.; Patra, G.; Muier, C.; Los, T.; Ivanova, .; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letes-

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A/Title: The genome sequence of the facultative intracellular pathogen *Brucella meliten*.

A/Reference number: AD3252; PMID:11756688

A/Accession: AF3382

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-224 <KUR>

A/Cross-references: UNIPROT:Q8YGM1; UNIPROT:Q8G102; GB:AE008917; PIDN:AAU52225.1; PID:91

A/Experimental source: strain 16M

C/Genetic:

A/Map position: I

Query Match 88.6%; Score 39; DB 2; Length 224;
Best Local Similarity 85.7%; Pred. No. 6.5;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
 Db 33 PHPOFG 39

RESULT 3

B97580
 hypothetical protein AGR_C_3351 [imported] - Agrobacterium tumefaciens (strain C58, Cere
 C/Species: Agrobacterium tumefaciens
 C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 09-Jul-2004
 C/Accession: B97580
 R/Gooder, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldman,
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;
 Science 294, 2323-2328, 2001
 A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
 A/Reference number: A97359; MUID:21608551; PMID:11743194
 A/Accession: B97580
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-225 <KUR>
 A/Cross-references: UNIPROT:Q8UED4; GB:AE007869; PIDN:AAK87595.1; PID:GL15156937; GSPDB:G
 C/Genetics: AGR_C_3351
 A/Map position: circular chromosome

Query Match 88.6%; Score 39; DB 2; Length 225;
 Best Local Similarity 85.7%; Pred. No. 6,6;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
 Db 33 PHPOFG 39

RESULT 4

AH2800
 conserved hypothetical protein Atu1826 [imported] - Agrobacterium tumefaciens (strain C5
 C/Species: Agrobacterium tumefaciens
 C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
 C/Accession: AH2800
 R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I
 etage, G.; Gillet, W.; Grant, C.; Gentner, D.; Kutyavain, T.; Levy, R.; Li, M.; McClell
 i; Karp, P.; Romberg, P.; Zhang, S.
 Science 294, 2317-2323, 2001
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
 ster, E.W.
 A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A/Reference number: AB2577; MUID:21608550; PMID:11743193
 A/Accession: AH2800
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-225 <KUR>
 A/Cross-references: UNIPROT:Q8UED4; GB:AE008668; PIDN:AAI42822.1; PID:GL17740269; GSPDB:G
 A/Experimental source: strain C58 (Dupont)
 C/Genetics: AGR_C_3351
 A/Map position: circular chromosome

Query Match 88.6%; Score 39; DB 2; Length 225;
 Best Local Similarity 85.7%; Pred. No. 6,6;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
 Db 33 PHPOFG 39

RESULT 5

S55900
 DNAU-like protein homolog - fission yeast (Schizosaccharomyces pombe)
 C/Species: Schizosaccharomyces pombe
 C/Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 09-Jul-2004
 C/Accession: S55900

R/Parf, S.K.; Chon, S.K.; Yoo, H.S.
 Biochim. Biophys. Acta 1262, 87-90, 1995
 A/Title: A cDNA of Schizosaccharomyces pombe encoding a homologue of DnaJ-like protein.
 A/Reference number: S55900; MUID:95290501; PMID:7772606
 A/Accession: S55900
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-379 <PAR>
 A/Cross-references: UNIPROT:Q09912; EMBL:U37753; NID:G576932; PIDN:AAA74732.1; PID:G9532
 C/Genetics: AGR_C_3351
 A/Map position: circular chromosome

Query Match 88.6%; Score 39; DB 2; Length 379;
 Best Local Similarity 75.0%; Pred. No. 11;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPQFG 8
 Db 187 RSHPSFG 194

RESULT 6

T41633
 psi protein - fission yeast (Schizosaccharomyces pombe)
 C/Species: Schizosaccharomyces pombe
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
 C/Accession: T41633
 R/McDougal, R.M.; Rajadream, M.A.; Barrell, B.G.; Ramsperger, U.; Pohl, T.
 submitted to the EMBL Data Library, August 1999
 A/Reference number: Z22005
 A/Accession: T41633
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-379 <MCD>
 A/Cross-references: UNIPROT:Q09912; EMBL:AL109850; PIDN:CAB52880.1; GSPDB:GN00068; SPDB:
 A/Experimental source: strain 972h-, cosmid c830
 C/Genetics: SPDB:SPCC830.07C
 A/Map position: 3
 A/Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 F/6-68/domain: dnaJ amino-terminal homology <DND>

Query Match 88.6%; Score 39; DB 2; Length 379;
 Best Local Similarity 75.0%; Pred. No. 11;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPQFG 8
 Db 187 RSHPSFG 194

RESULT 7

S08156
 gene T protein - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004
 C/Accession: S08156
 R/Herrmann, B.G.; Labelle, S.; Pouetka, A.; King, T.R.; Lehrach, H.
 Nature 343, 617-622, 1990
 A/Title: Cloning of the T gene required in mesoderm formation in the mouse.
 A/Reference number: S08156; MUID:90158787; PMID:2154654
 A/Accession: S08156
 A/Status: not compared with conceptual translation
 A/Molecule type: mRNA
 A/Residues: 1-436 <HER>
 A/Cross-references: UNIPROT:P20293; GB:X51683; NID:G55053; PIDN:CAA35985.1; PID:G55054
 C/Genetics: AGR_C_3351
 A/Map position: 17
 A/Superfamily: mouse gene T protein; T-box homology
 C/Keywords: DNA binding; transcription regulation

F:51-227/Domain: T-box homology <TBX>

Query Match 88.6%; Score 39; DB 1; Length 436;
Best Local Similarity 85.7%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
:|||||
Db 260 SHPQFG 266

RESULT 8

H66216 protein T23G18.16 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004

C:Accession: H66216

R:Thellogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Xu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A66141; MUID:21016719; PMID:11130712

A:Accession: H66216

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-514 <STO>

A:Cross-references: UNIPROT:Q9SGD0; GB:AE005172; NID:G6579199; PIND:AAF18242.1; GSPDB:GN

C:Genetics: T23G18.1d

A:Map position: 1

Query Match 88.6%; Score 39; DB 2; Length 512;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
:|||||
Db 112 PHPQFG 118

RESULT 9

B69256 conserved hypothetical protein AF0050 - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 09-Jul-2004

C:Accession: B69256

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson,
; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.
Glock, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 350, 364-370, 1997

A:Authors: Ustebach, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.
Smith, H.O.; Wose, C.R.; Venter, J.C.
A>Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo

A:Reference number: A69250; MUID:98049343; PMID:9389475

A:Accession: B69256

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-399 <KLE>

A:Cross-references: UNIPROT:Q30186; GB:AE001103; GB:AE000782; NID:G2689426; PIND:AAF9118

Query Match 84.1%; Score 37; DB 2; Length 399;
Best Local Similarity 75.0%; Pred. No. 28;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RXHPQFG 8
:|||||
Db 70 RSNPQFG 77

RESULT 10

B87712

conserved hypothetical protein CC3732 [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004

C:Accession: B87712

R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DebRoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A>Title: Complete Genome Sequence of Caulobacter crescentus.

A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: B87712

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-309 <STO>

A:Cross-references: UNIPROT:Q9A234; GB:AE005673; NID:G13425504; PIND:AAK25694.1; GSPDB:CC

C:Genetics: CC3732

Query Match 81.8%; Score 36; DB 2; Length 309;
Best Local Similarity 75.0%; Pred. No. 33;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RXHPQFG 8
:|||||
Db 228 RVQPOFG 235

RESULT 11

B82358

alkaline serine proteinase VC0157 [imported] - Vibrio cholerae (strain N16961 serogroup

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004

C:Accession: B82358

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.,
chardson, D.; Ermolaeva, M.D.; Vamathavan, J.; Bass, S.; Qin, H.; Dregol, I.; Sellers, F.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000

A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: B82358

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-535 <HEI>

A:Cross-references: UNIPROT:Q9KV18; GB:AE004105; GB:AE003852; NID:G9654551; PIND:AAF9333

A:Experimental source: serogroup O1, strain N16961, biotype El Tor

C:Genetics:

A:Gene: VC0157

A:Map position: 1

C:Superfamily: subtilisin; subtilisin homology

Query Match 81.8%; Score 36; DB 2; Length 535;
Best Local Similarity 71.4%; Pred. No. 57;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
:|||||
Db 184 NHPQFG 190

RESULT 12

S18106

type II site-specific deoxyribonuclease (EC 3.1.21.4) Abri - Azospirillum brasilense

N:Alternate names: restriction endonuclease Abri

C:Species: Azospirillum brasilense

C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 21-Aug-1998

C:Accession: S18106

R:Schwabe, G.; Helke, A.; Klinkmuller, W.

submitted to the EMBL Data Library, October 1991

A:Description: Abri restriction modification system from Azospirillum brasilense, molec

A:Reference number: S18106
 A:Accession: S18106
 A:Molecule type: DNA
 A:Residues: 1-222 <SCH-
 A:Cross-references: EMBL:X62690
 C:Superfamily: type II site-specific deoxyribonuclease Paer71
 C:Keywords: hydrolase

Query Match 79.5%; Score 35; DB 2; Length 222;
 Best Local Similarity 62.5%; Pred. No. 37;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPQFG 8
 |||:|
 Db 178 RHPFYGG 185

RESULT 13
 T12462
 hypothetical protein DKFZ5641122.1 - human (fragment)
 C:Species: Homo sapiens (man)
 C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 09-Jul-2004
 C:Accession: T12462
 R:Duesterhoeft, A.; Lauber, J.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, June 1999
 A:Reference number: Z17522
 A:Accession: T12462
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-288 <DUS-
 A:Cross-references: UNIPROT:Q9Y4U1; EMBL:AL080062
 A:Experimental source: fetal brain; clone DKFZ5641122
 C:Genetics:
 A:Note: DKFZ5641122.1

Query Match 79.5%; Score 35; DB 2; Length 288;
 Best Local Similarity 71.4%; Pred. No. 47;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
 |||:|
 Db 156 IHPRFGG 162

RESULT 14
 G84105
 transcription regulator BH3647 [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 09-Jul-2004
 C:Accession: G84105
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Maui, N.; Fuji, F.; Hira
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: G84105
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-359 <STO-
 A:Cross-references: UNIPROT:Q9K659; GB:AP001519; GB:BA000004; NID:G10176109; PIDN:BAB073
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH3647
 C:Superfamily: Bacillus subtilis probable transcription regulator ywhU

Query Match 79.5%; Score 35; DB 2; Length 359;
 Best Local Similarity 75.0%; Pred. No. 59;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPQFG 8
 |||:|
 Db 213 RHPQGG 220

RESULT 15

hypothetical protein At2g18880 [imported] - Arabidopsis thaliana
 T01616
 N:Alternate names: hypothetical protein F19F24.8
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 09-Jul-2004
 C:Accession: T01616; G84569
 R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
 submitted to the EMBL Data Library, April 1998
 A:Description: Arabidopsis thaliana chromosome II BAC F19F24 genomic sequence.
 A:Reference number: Z14153
 A:Accession: T01616
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-417 <ROU>
 A:Cross-references: UNIPROT:O64618; EMBL:AC003673; NID:G3004543; PID:G3004551
 A:Experimental source: cultivar Columbia
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Umayam, L.; Tallon, L.
 euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: G84569
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-417 <STO>
 A:Cross-references: GB:AE002093; NID:G3004551; PIDN:AAC09024.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: F19F24.8; At2g18880
 A:Map position: 2
 A:Introns: 103/3; 200/1

Query Match 79.5%; Score 35; DB 2; Length 417;
 Best Local Similarity 62.5%; Pred. No. 69;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPQFG 8
 |||:|
 Db 396 RHPPIGG 403

Search completed: March 2, 2005, 12:28:53
 Job time : 7.11382 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 29.7236 Seconds
(without alignments)
137.824 Million cell updates/sec

Title: SEQ7
Perfect score: 44
Sequence: 1 rxhpqfsg 8

Scoring table: BLOSUM62DX
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : UniProt 03: *
1: uniprot_sprot: *
2: uniprot_trembl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	88.6	67	2	Q8SZ12 drosophila
2	39	88.6	92	2	Q02737 sus scrofa
3	39	88.6	212	2	Q97669 oryctolagus
4	39	88.6	224	2	Q8Y6W1 bruceella me
5	39	88.6	224	2	Q8G102 bruceella su
6	39	88.6	225	2	Q92PK7 rhizobium m
7	39	88.6	225	2	Q8UED4 agrobacteri
8	39	88.6	226	2	Q6FZW3 bartonella
9	39	88.6	226	2	Q6G3B7 bartonella
10	39	88.6	228	2	Q96NR9 rhizobium 1
11	39	88.6	330	2	Q96UF7 metarhizium
12	39	88.6	379	1	PSII SCHPO
13	39	88.6	385	2	Q96UF9 metarhizium
14	39	88.6	385	2	Q96UF9 metarhizium
15	39	88.6	386	2	Q96UF9 metarhizium
16	39	88.6	435	1	BRAC HUMAN
17	39	88.6	435	1	BRAC HUMAN
18	39	88.6	436	1	BRAC MOUSE
19	39	88.6	512	2	Q9SGD0
20	39	88.6	512	2	Q9SGD0
21	39	88.6	647	2	Q8ZJP5 streptomyces
22	39	88.6	1074	2	Q9WJW1 drosophila
23	39	88.6	297	2	Q7NT66 chromobacter
24	39	88.6	356	2	Q72CV2 deaullifovdr
25	39	88.6	406	2	Q8ZK33 salmoneilla
26	39	88.6	407	2	Q8FAD9 escherichia
27	39	88.6	542	2	Q8N7C2 homo sapien
28	39	88.6	542	2	Q8N7C2 homo sapien
29	37	84.1	198	1	Q9L240 streptomyces
30	36	81.8	181	2	Q71DA7 archaebact
31	36	81.8	181	2	Q9GP12 drosophila

32	36	81.8	181	2	Q9VEH5	Q9VEH5 drosophila
33	36	81.8	232	2	Q73H66	Q73H66 wolbachia p
34	36	81.8	260	2	Q750S3	Q750S3 ashbya goss
35	36	81.8	294	2	Q7UVW5	Q7UVW5 rhodospirillum
36	36	81.8	309	1	KERS_WOLSU	Q7MBJ0 wolbachia s
37	36	81.8	309	2	Q9A234	Q9A234 caulobacter
38	36	81.8	311	2	Q6AF78	Q6AF78 leifsonia x
39	36	81.8	323	2	Q6TR93	Q6TR93 metarhizium
40	36	81.8	348	2	Q8SR57	Q8SR57 encyrtid
41	36	81.8	356	2	Q8CNS6	Q8CNS6 staphylococ
42	36	81.8	390	2	Q96UF8	Q96UF8 metarhizium
43	36	81.8	390	2	Q9UDR5	Q9UDR5 metarhizium
44	36	81.8	400	1	TRUD METKA	Q8XJ7 methanopyru
45	36	81.8	440	2	Q6C012	Q6C012 yarrowia li

ALIGNMENTS

RESULT 1	ID	Q8SZ12	PRELIMINARY;	PRT;	67 AA.
AC	Q8SZ12				
DT	01-JUN-2002	(TREMblrel. 21, Created)			
DT	01-JUN-2002	(TREMblrel. 21, Last sequence update)			
DT	01-OCT-2002	(TREMblrel. 22, Last annotation update)			
DE	LD08951P.				
GN	ORFName=CG4532;				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;				
OC	Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Berkley;				
RA	Stapleton M., Brooksstein P., Hong L., Agbayani A., Carlson J.,				
RA	Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,				
RA	George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,				
RA	Miranda A., Mungall C.J., Nuncio J., Pacheco J., Paragas V., Park S.,				
RA	Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,				
RA	Celniker S.;				
RL	Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AY070876; AAL48498.1; -				
DR	Flybase; FBgn0029903; CG4532.				
SQ	SEQUENCE 67 AA; 7782 MW; F9C917BE81577B94 CRC64;				
Query Match					
Best Local Similarity	85.6%;	Score 39;	DB 2;	Length 67;	
Matches	6;	Conservative	1;	Mismatches	0;
				Indels	0;
				Gaps	0;
QY	2 XHPQFG 8				
DB	19 DHPQFG 25				
RESULT 2					
ID	Q02737	PRELIMINARY;	PRT;	92 AA.	
AC	Q02737				
DT	01-JUL-1997	(TREMblrel. 04, Created)			
DT	01-JUL-1997	(TREMblrel. 04, Last sequence update)			
DT	01-DEC-2001	(TREMblrel. 19, Last annotation update)			
DE	Brachyury (Fragment).				
OS	Sus scrofa (Pig).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.				
OX	NCBI_TaxID=9823;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RA	Yellich J.V., Pomp D., Geisert R.D.;				
RL	Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; U91519; AAB53317.1; -				

```

DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003700; F:transcription factor activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR002070; TF_Brachyury.
DR PRINTS; PR00938; BRACHYURY.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 92 AA; 10134 MW; 7AE235EB7B32AD7D CRC64;

Query Match 88.6%; Score 39; DB 2; Length 92;
Best Local Similarity 85.7%; Pred. No. 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 51 PHPQFG 57

RESULT 3
O97669 PRELIMINARY; PRT; 212 AA.
ID O97669;
AC O97669;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Brachyury (T) protein (Fragment).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=New Zealand White;
RC MEDLINE=21969334; PubMed=11973268;
RX Vietham C., Stortz C., Mitchell S.A., Blum M.;
RA "Low proliferative and high migratory activity in the area of
RT Brachyury expressing mesoderm progenitor cells in the gastrulating
RT rabbit embryo.";
RL Development 129:2355-2365(2002).
DR HSP; P24781; IXR.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003700; F:transcription factor activity; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR008967; P53 like DNA_bnd.
DR InterPro: IPR002070; TF_Brachyury.
DR InterPro: IPR001699; TF_T-box.
DR Pfam; PF00907; T-box_1
DR PRINTS; PR00938; BRACHYURY.
DR PRINTS; PR00937; TBOX.
DR SMART; SM00425; TBOX_1.
DR PROSITE; PSS0252; TBOX_3; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 212 AA; 22864 MW; CB7F1D39EA372B4C CRC64;

Query Match 88.6%; Score 39; DB 2; Length 212;
Best Local Similarity 85.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 114 PHPQFG 120

RESULT 4
O8YGM1 PRELIMINARY; PRT; 224 AA.
ID O8YGM1;
AC O8YGM1;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE ALPHA/BETA HYDROLASE.

```

```

GN OrderedLocustNames=EMBL1044;
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales.
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=16M / ATCC 23456 / Biotype 1;
RC MEDLINE=20020109; PubMed=11756688; DOI=10.1073/pnas.221575398;
RX Delvecchio V.G., Kaputal V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykdis A., Reznik G.,
RA Jabloncki L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elser P.H., Hagius S., O'Callaghan D., Teleson J.,
RA Haselkorn R., Kyrides N.C., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009544; AL52225.1; -.
DR PIR; AF3382; AF3382.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR Complete proteome.
KW SEQUENCE 224 AA; 24999 MW; 882EFP312441586 CRC64;

Query Match 88.6%; Score 39; DB 2; Length 224;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 33 PHPQFG 39

RESULT 5
O8G102 PRELIMINARY; PRT; 224 AA.
ID O8G102;
AC O8G102;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=BR0929;
OS Brucella suis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales.
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29461;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=1330 / Biovar 1;
RX MEDLINE=22247741; PubMed=12271122; DOI=10.1073/pnas.192319099;
RA Paulsen I.T., Sehadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
RA Read T.D., Dodson R.J., Umayam L.A., Brinkac L.M., Beaman M.J.,
RA Daugherty S.C., Debey R.T., Durkin A.S., Kolonay J.F., Madupu R.,
RA Nelson W.C., Ayodeji B., Kraut M., Shetty J., Malek J.A.,
RA Van Aken S.E., Riedmiller S., Tettelin H., Gill S.R., White O.,
RA Salzberg S.L., Hoover D.L., Lindler L.E., Halling S.M., Boyle S.M.,
RA Fraser C.M.;
RT "The Brucella suis genome reveals fundamental similarities between
RT animal and plant pathogens and symbionts."
RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
DR EMBL; AE014395; AAN29855.1; -.
DR PIR; AF3382; AF3382.
DR TIGR; BR0929; -.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR InterPro: IPR000379; Ser_estrs.
DR Complete proteome.
KW SEQUENCE 224 AA; 24999 MW; 882EFP312441586 CRC64;

Query Match 88.6%; Score 39; DB 2; Length 224;
Best Local Similarity 85.7%; Pred. No. 23;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8

```